(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number
WO 02/40541 A2

(51) International Patent Classification[7]: C07K 14/705

(21) International Application Number: PCT/US01/46055

(22) International Filing Date: 25 October 2001 (25.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/243,989	27 October 2000 (27.10.2000)	US
60/245,904	3 November 2000 (03.11.2000)	US
60/247,673	9 November 2000 (09.11.2000)	US
60/249,661	17 November 2000 (17.11.2000)	US
60/252,232	20 November 2000 (20.11.2000)	US
60/250,790	1 December 2000 (01.12.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **SANJANWALA, Madhusudan** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way, #1, Menlo Park, CA 94025 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **KHAN, Farrah, A.** [IN/US]; 9445 Harrison Street, Des Plaines, IL 60016 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue, #23, Mountain View, CA 94043 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive, #2, Cupertino, CA 95014 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **DAS, Debopriya** [IN/US]; 1267 Parkington Avenue, Sunnyvale, CA 94087 (US). **RAUMANN, Brigette, E.** [US/US]; 5801 South Dorchester Avenue #3B, Chicago, IL 60637 (US). **POLICKY, Jennifer, L.** [US/US]; 1511 Court Jarvis Court, San Jose, CA 95118 (US). **KEARNEY, Liam** [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.

WO 02/40541 A2

TRANSPORTERS AND ION CHANNELS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transporters and ion channels and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological, muscle, immunological and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including K^+, NH_4^+, P_i, SO_4^{2-}, sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous Na^+/K^+ ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging

techniques and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc. Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., biotin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel syndrome, and non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H^+-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H^+-linked monocarboxylate transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K_m values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na^+-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective transporters for organic cations and organic anions in organs including the

kidney, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH (Poole, R.C. and A.P. Halestrap (1993) Am. J. Physiol. 264:C761-C782; Price, N.T. et al. (1998) Biochem. J. 329:321-328; and Martinelle, K. and I. Haggstrom (1993) J. Biotechnol. 30:339-350).

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids, peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes. When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) Meth. Enzymol. 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) J. Med. Genet. 23:99-106).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty acid transport appears to occur via a high affinity, low capacity protein-mediated transport process.

Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments, is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et al. (1998) J. Biol. Chem. 273:27420-27429).

The lipocalin superfamily constitutes a phylogenetically conserved group of more than forty proteins that function as extracellular ligand-binding proteins which bind and transport small hydrophobic molecules. Members of this family function as carriers of retinoids, odorants, chromophores, pheromones, allergens, and sterols, and in a variety of processes including nutrient transport, cell growth regulation, immune response, and prostaglandin synthesis. A subset of these proteins may be multifunctional, serving as either a biosynthetic enzyme or as a specific enzyme inhibitor. (Tanaka, T. et al. (1997) J. Biol. Chem. 272:15789-15795; and van't Hof, W. et al. (1997) J. Biol. Chem. 272:1837-1841.)

Members of the lipocalin family display unusually low levels of overall sequence conservation. Pairwise sequence identity often falls below 20%. Sequence similarity between family members is limited to conserved cysteines which form disulfide bonds and three motifs which form a juxtaposed cluster that functions as a target cell recognition site. The lipocalins share an eight stranded, anti-parallel beta-sheet which folds back on itself to form a continuously hydrogen-bonded beta-barrel. The pocket formed by the barrel functions as an internal ligand binding site. Seven loops (L1 to L7) form short beta-hairpins, except loop L1 which is a large omega loop that forms a lid to partially close the internal ligand-binding site (Flower (1996) Biochem. J. 318:1-14).

Lipocalins are important transport molecules. Each lipocalin associates with a particular ligand and delivers that ligand to appropriate target sites within the organism. Retinol-binding protein (RBP), one of the best characterized lipocalins, transports retinol from stores within the liver to target tissues. Apolipoprotein D (apo D), a component of high density lipoproteins (HDLs) and low density lipoproteins (LDLs), functions in the targeted collection and delivery of cholesterol throughout the body. Lipocalins are also involved in cell regulatory processes. Apo D, which is identical to gross-cystic-disease-fluid protein (GCDFP)-24, is a progesterone/pregnenolone-binding protein expressed at high levels in breast cyst fluid. Secretion of apo D in certain human breast cancer cell lines is accompanied by reduced cell proliferation and progression of cells to a more differentiated phenotype. Similarly, apo D and another lipocalin, α_1-acid glycoprotein (AGP), are involved in nerve cell regeneration. AGP is also involved in anti-inflammatory and immunosuppressive activities. AGP is one of the positive acute-phase proteins (APP); circulating levels of AGP increase in response to stress and inflammatory stimulation. AGP accumulates at sites of inflammation where it inhibits platelet and neutrophil activation and inhibits phagocytosis. The

immunomodulatory properties of AGP are due to glycosylation. AGP is 40% carbohydrate, making it unusually acidic and soluble. The glycosylation pattern of AGP changes during acute-phase response, and deglycosylated AGP has no immunosuppressive activity (Flower (1994) FEBS Lett. 354:7-11; Flower (1996) supra).

The lipocalin superfamily also includes several animal allergens, including the mouse major urinary protein (mMUP), the rat α-2-microgloobulin (rA2U), the bovine β-lactoglobulin (βlg), the cockroach allergen (Bla g4), bovine dander allergen (Bos d2), and the major horse allergen, designated *Equus caballus* allergen 1 (Equ c1). Equ c1 is a powerful allergen responsible for about 80% of anti-horse IgE antibody response in patients who are chronically exposed to horse allergens. It appears that lipocalins may contain a common structure that is able to induce the IgE response (Gregoire, C. et al., (1996) J. Biol. Chem. 271:32951-32959).

Lipocalins are used as diagnostic and prognostic markers in a variety of disease states. The plasma level of AGP is monitored during pregnancy and in diagnosis and prognosis of conditions including cancer chemotherapy, renal disfunction, myocardial infarction, arthritis, and multiple sclerosis. RBP is used clinically as a marker of tubular reabsorption in the kidney, and apo D is a marker in gross cystic breast disease (Flower (1996) supra). Additionally, the use of lipocalin animal allergens may help in the diagnosis of allergic reactions to horses (Gregoire supra), pigs, cockroaches, mice and rats.

Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) *275000 Graves Disease).

This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) J. Int. Med. 245:637-642).

I n Channels

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

Ion Transporters

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient. These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion transporters, including Na^+-K^+ ATPase, Ca^{2+}-ATPase, and H^+-ATPase, are activated by a phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of Na^+ and Ca^{2+} are low and cytosolic concentration of K^+ is high. The vacuolar (V) class of ion transporters includes H^+ pumps on intracellular organelles, such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class consists of H^+ pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate (P_i).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) Curr. Opin. Cell Biol. 11:517-522). The V-ATPases are composed of two functional domains: the V_1 domain, a peripheral complex responsible for ATP hydrolysis; and the V_0 domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and evolutionarily related to the V-ATPases. The F-ATPase F_0 domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase V_0 domain contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgac, M. (1999) J. Biol. Chem. 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and

gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of Na^+ down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of Ca^{2+} out of the cell with transport of Na^+ into the cell (antiport).

Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction, fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated channels open their pores in response to mechanical stress; voltage-gated channels (e.g., Na^+, K^+, Ca^{2+}, and Cl^- channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g., acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post translational modifications, such as phosphorylation.

Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or stress on the cell membrane and conduct both Ca^{2+} and Na^+ (Suzuki, M. et al. (1999) J. Biol. Chem. 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and carboxy termini. In the Na^+ and Ca^{2+} subfamilies, this domain is repeated four times, while in the K^+ channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of K^+ channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated Na^+ and K^+ channels are necessary for the function of electrically excitable cells, such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to Na^+

and K^+ ions. Depolarization of the membrane beyond the threshold level opens voltage-gated Na^+ channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more voltage-gated Na^+ channels, which propagates the depolarization down the length of the cell. Depolarization also opens voltage-gated potassium channels. Consequently, potassium ions flow outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's N-terminus, which acts as a plug that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated Na^+ channels are heterotrimeric complexes composed of a 260 kDa pore-forming α subunit that associates with two smaller auxiliary subunits, β1 and β2. The β2 subunit is a integral membrane glycoprotein that contains an extracellular Ig domain, and its association with α and β1 subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) Cell 83:433-442).

Non voltage-gated Na^+ channels include the members of the amiloride-sensitive Na^+ channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial Na^+ channel (ENaC) involved in Na^+ reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized H^+-gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric Na^+-permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from C. elegans. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) Curr. Opin. Neurobiol. 8:418-424; Eglen, R.M. et al. (1999) Trends Pharmacol. Sci. 20:337-342).

K^+ channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as Ca^{2+} and cAMP. In non-excitable tissue, K^+ channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes, K^+ channels are responsible for setting the resting membrane

potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a Na^+-K^+ pump and ion channels that provide the redistribution of Na^+, K^+, and Cl^-. The pump actively transports Na^+ out of the cell and K^+ into the cell in a 3:2 ratio. Ion channels in the plasma membrane allow K^+ and Cl^- to flow by passive diffusion. Because of the high negative charge within the cytosol, Cl^- flows out of the cell. The flow of K^+ is balanced by an electromotive force pulling K^+ into the cell, and a K^+ concentration gradient pushing K^+ out of the cell. Thus, the resting membrane potential is primarily regulated by K^+ flow (Salkoff, L. and T. Jegla (1995) Neuron 15:489-492).

Potassium channel subunits of the Shaker-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form functional K channels. These pore-forming subunits also associate with various cytoplasmic β subunits that alter channel inactivation kinetics. The Shaker-like channel family includes the voltage-gated K^+ channels as well as the delayed rectifier type channels such as the human ether-a-go-go related gene (HERG) associated with long QT, a cardiac dysrythmia syndrome (Curran, M.E. (1998) Curr. Opin. Biotechnol. 9:565-572; Kaczorowski, G.J. and M.L. Garcia (1999) Curr. Opin. Chem. Biol. 3:448-458).

A second superfamily of K^+ channels is composed of the inward rectifying channels (Kir). Kir channels have the property of preferentially conducting K^+ currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated K^+ channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac pacemaker activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) Curr. Opin. Neurobiol. 5:268-277; Curran, supra).

The recently recognized TWIK K^+ channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential in a large set of cell types (Duprat, F. et al. (1997) EMBO J 16:5464-5471).

The voltage-gated Ca^{2+} channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type Ca^{2+} channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated Ca^{2+} channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of

three subunits. The α_1 subunit forms the membrane pore and voltage sensor, while the $\alpha_2\delta$ and β subunits modulate the voltage-dependence, gating properties, and the current amplitude of the channel. These subunits are encoded by at least six α_1, one $\alpha_2\delta$, and four β genes. A fourth subunit, γ, has been identified in skeletal muscle (Walker, D. et al. (1998) J. Biol. Chem. 273:2361-2367; McCleskey, E.W. (1994) Curr. Opin. Neurobiol. 4:304-312).

The high-voltage-activated Ca^{2+} channels that have been characterized biochemically include complexes of a pore-forming alpha1 subunit of approximately 190-250 kDa; a transmembrane complex of alpha2 and delta subunits; an intracellular beta subunit; and in some cases a transmembrane gamma subunit. A variety of alpha1 subunits, alpha2delta complexes, beta subunits, and gamma subunits are known. The Cav1 family of alpha1 subunits conduct L-type Ca^{2+} currents, which initiate muscle contraction, endocrine secretion, and gene transcription, and are regulated primarily by second messenger-activated protein phosphorylation pathways. The Cav2 family of alpha1 subunits conduct N-type, P/Q-type, and R-type Ca^{2+} currents, which initiate rapid synaptic transmission and are regulated primarily by direct interaction with G proteins and SNARE proteins and secondarily by protein phosphorylation. The Cav3 family of alpha1 subunits conduct T-type Ca^{2+} currents, which are activated and inactivated more rapidly and at more negative membrane potentials than other Ca^{2+} current types. The distinct structures and patterns of regulation of these three families of Ca^{2+} channels provide an array of Ca^{2+} entry pathways in response to changes in membrane potential and a range of possibilities for regulation of Ca^{2+} entry by second messenger pathways and interacting proteins (Catterall, W.A. (2000) Annu. Rev. Cell Dev. Biol. 16:521-555).

The alpha-2 subunit of the voltage-gated Ca^{2+}-channel may include one or more Cache domains. An extracellular Cache domain may be fused to an intracellular catalytic domain, such as the histidine kinase, PP2C phosphatase, GGDEF (a predicted diguanylate cyclase), HD-GYP (a predicted phosphodiesterase) or adenylyl cyclase domain, or to a noncatalytic domain, like the methyl-accepting, DNA-binding winged helix-turn-helix, GAF, PAS or HAMP (a domain found in istidine kinases, denylyl cyclases, ethyl-binding proteins and phosphatases). Small molecules are bound via the Cache domain and this signal is converted into diverse outputs depending on the intracellular domains (Anantharaman, V. and Aravind, L.(2000) Trends Biochem. Sci. 25:535-537).

The transient receptor family (Trp) of calcium ion channels are thought to mediate capacitative calcium entry (CCE). CCE is the Ca^{2+} influx into cells to resupply Ca^{2+} stores depleted by the action of inositol triphosphate (IP3) and other agents in response to numerous hormones and growth factors. Trp and Trp-like were first cloned from Drosophila and have similarity to voltage gated Ca^{2+} channels in the S3 through S6 regions. This suggests that Trp and/or related proteins may form mammalian CCE channels (Zhu, X. et al. (1996) Cell 85:661-671; Boulay, G. et al. (1997) J. Biol. Chem. 272:29672-29680). Melastatin is a gene isolated in both the mouse and human, whose

expression in melanoma cells is inversely correlated with melanoma aggressiveness in vivo. The human cDNA transcript corresponds to a 1533-amino acid protein having homology to members of the Trp family. It has been proposed that the combined use of malastatin mRNA expression status and tumor thickness might allow for the determination of subgroups of patients at both low and high risk for developing metastatic disease (Duncan, L.M. et al (2001) J. Clin. Oncol. 19:568-576).

Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and organelle pH. In secretory epithelial cells, Cl^- enters the cell across a basolateral membrane through an Na^+, K^+/Cl^- cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of Cl^- from the apical surface, in response to hormonal stimulation, leads to flow of Na^+ and water into the secretory lumen. The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) J. Exp. Biol. 172:245-266).

The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume, membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) Curr. Opin. Neurobiol. 6:303-310).

Ligand-gated channels open their pores when an extracellular or intracellular mediator binds to the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of Na^+ and produces the initial localized depolarization that activates the voltage-gated channels and starts the action potential. Chloride channels open in response to inhibitory neurotransmitters, such as γ-aminobutyric acid (GABA) and glycine, leading to hyperpolarization of the membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, supra). Amino acids in the second transmembrane domain appear to be important

in determining channel permeation and selectivity (Sather, W.A. et al. (1994) Curr. Opin. Neurobiol. 4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated K^+ channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens K^+ channels to modulate the magnitude of the action potential (Ishi et al., supra). The large conductance (BK) channel has been purified from brain and its subunit composition determined. The α subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated K^+ channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The β subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, supra; Vergara, C. et al. (1998) Curr. Opin. Neurobiol. 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated Na^+ channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also represent a major pathway for Ca^{2+} entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an α subunit which can form functional homomeric channels, and a β subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated K^+ channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) Curr. Opin. Neurobiol. 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of intracellular signalling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the Gβγ subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) Curr. Opin. Cell. Biol. 11:503-508). Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Bredt (1998) Cell 93:495-498).

Disease Correlati n

The etiology of numerous human diseases and disorders can be attributed to defects in the

transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff, W.G. (1996) Exp. Nephrol. 4:253-262; Talente, G.M. et al. (1994) Ann. Intern. Med. 120:218-226; and Chillon, M. et al. (1995) New Engl. J. Med. 332:1475-1480).

Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrythmia disorders such as the long QT syndromes and idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper, E.C. and L.Y. Jan (1998) Proc. Natl. Acad. Sci. USA 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and I.E. Scheffer (1999) Curr. Opin. Neurology 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) Curr. Opin. Neurobiol. 9:274-280; Cooper, supra).

Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) Adv. Pharmacol. 39:47-98). Various classes of ion channels also play an important role in the perception of pain, and thus are potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated Na^+ channels have been useful in the treatment of neuropathic pain (Eglen, supra).

Ion channels in the immune system have recently been suggested as targets for immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious

immunosuppressants (Cahalan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

The discovery of new transporters and ion channels, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological, muscle, immunological and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, transporters and ion channels, referred to collectively as "TRICH" and individually as "TRICH-1," "TRICH-2," "TRICH-3," "TRICH-4," "TRICH-5," "TRICH-6," "TRICH-7," "TRICH-8," "TRICH-9," "TRICH-10," "TRICH-11," "TRICH-12," "TRICH-13," "TRICH-14," "TRICH-15," "TRICH-16," "TRICH-17," "TRICH-18," "TRICH-19," and "TRICH-20." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-20. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:21-40.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least

90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide

sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a

naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino

acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in

an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TRICH" refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRICH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRICH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRICH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine.

Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TRICH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include

deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide

or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TRICH or fragments of TRICH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of

the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of TRICH or the polynucleotide encoding TRICH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:21-40 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:21-40, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:21-40 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish

SEQ ID NO:21-40 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:21-40 and the region of SEQ ID NO:21-40 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-20 is encoded by a fragment of SEQ ID NO:21-40. A fragment of SEQ ID NO:1-20 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-20. For example, a fragment of SEQ ID NO:1-20 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-20. The precise length of a fragment of SEQ ID NO:1-20 and the region of SEQ ID NO:1-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with

other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of

polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill

in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TRICH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acid sequences encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target

polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and

polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular

structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection,

transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, immunological and cell proliferative disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide

sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homologs along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transporters and ion channels. For example, SEQ ID NO:5 is 61% identical to Drosophila sodium-hydrogen exchanger NHE1 (GenBank ID g4894991) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.0e-139, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a sodium/hydrogen exchanger family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provides further corroborative evidence that SEQ ID NO:5 is a sodium/hydrogen exchanger. In an alternative example, SEQ ID NO:6 is about 50% identical to human citrin, the adult-onset type II citrullinemia protein, (GenBank ID g5052319) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is

6.0e-51, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains mitochondrial carrier protein domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:6 is a mitochondrial carrier protein. In an alternative example, SEQ ID NO:7 is 27% identical to Synechocystis sp. melibiose carrier protein (GenBank ID g1653342) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.8e-16, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Additional BLAST data from DOMO and PRODOM analyses provide further corroborative evidence that SEQ ID NO:7 is a symporter protein. In an alternative example, SEQ ID NO:9 is 26% identical to an Arabidopsis ABC transporter (GenBank ID g4262239) and is 99% identical, from residue M1 to residue W374, to human sterolin-2 (GenBank ID g15146444) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability scores are 4.1e-25 and 0.0 respectively, which indicate the probabilities of obtaining the observed polypeptide sequence alignments by chance. SEQ ID NO:9 contains two transmembrane domains as determined by hidden Markov model (HMM) analysis, as well as a white/scarlet ABC transporter domain. (See Table 3.) These data provide further corroborative evidence that SEQ ID NO:9 is an ABC transporter. In an alternative example, SEQ ID NO:12 is 93% identical to rat neuronal glutamine transporter (GenBank ID g6978016) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.4e-239, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a transmembrane amino acid transporter domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) These data provide corroborative evidence that SEQ ID NO:12 is an amino acid transporter protein. In an alternative example, SEQ ID NO:14 is 52% identical to mouse multidrug resistance protein (GenBank ID g387426) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 also contains an ABC transporter domain and an ABC transporter transmembrane region domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:14 is a multidrug resistance ABC transporter. In an alternative example, SEQ ID NO:18 is 41% identical to Arabidopsis putative membrane transporter (GenBank ID g2289003) and is 99% identical, from

residue M20 to residue E648, to human proton myo-inositol transporter (GenBank ID g15211933) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability scores are 1.4e-94 and 0.0 respectively, which indicate the probabilities of obtaining the observed polypeptide sequence alignments by chance. SEQ ID NO:18 also contains a sugar (and other) transporter domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:18 is a sugar transporter. SEQ ID NO:1-4, SEQ ID NO:8, SEQ ID NO:10-11, SEQ ID NO:13, SEQ ID NO:15-17, and SEQ ID NO:19-20 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-20 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:21-40 or that distinguish between SEQ ID NO:21-40 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6122382H1 is the identification number of an Incyte cDNA sequence, and BRAHNON05 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 72008374V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2077361) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences

including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide

sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural characteristic of TRICH.

The invention also encompasses polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40, which encodes TRICH. The polynucleotide sequences of SEQ ID NO:21-40, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TRICH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:21-40. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding TRICH. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding TRICH, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding TRICH. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TRICH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRICH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRICH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:21-40 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is

automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze

the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TRICH.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable

manner.

In another embodiment, sequences encoding TRICH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TRICH, the nucleotide sequences encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TRICH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques,

and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TRICH can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TRICH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors

containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of TRICH. Transcription of sequences encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, sequences encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which

successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing sequences encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,

e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRICH may be designed to contain signal sequences which direct secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRICH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may

facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TRICH may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

TRICH of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TRICH. At least one and up to a plurality of test compounds may be screened for specific binding to TRICH. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TRICH binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TRICH of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TRICH may also be manipulated in vitro in ES cells derived from

human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. In addition, the expression of TRICH is closely associated with tumorous tissues such as spleen tumor tissue, esophageal tumor tissue, brain tumor tissue, and myxoma from atrium tissue; and normal tissues such as kidney, liver, nasal polyp, prostate, thyroid, umbilical cord blood, neuronal, digestive, uterine endometrial tissue, and normal brain tissue such as the tissues from striatum, globus pallidus, and posterior putamen. Therefore, TRICH appears to play a role in transport, neurological, muscle, immunological and cell proliferative disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy,

ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, glycogen storage disease, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, pseudohypoaldosteronism type 1, Liddle's syndrome, cystinuria, iminoglycinuria, Hartup disease, Fanconi disease, and Bartter syndrome; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, hemiplegic migraine, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, acid maltase

deficiency (AMD, also known as Pompe's disease), generalized myotonia, and myotonia congenita; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH. Examples of such disorders include, but are not limited to, those transport, neurological, muscle, immunological and cell proliferative disorders described above. In one aspect, an antibody which specifically binds

TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TRICH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma

technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TRICH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies

for TRICH. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TRICH, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res.

25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl.

Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver

polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity

(e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques

for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TRICH may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted

biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TRICH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated

with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRICH or closely related molecules may be used to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:21-40 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

Means for producing specific hybridization probes for DNAs encoding TRICH include the cloning of polynucleotide sequences encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease,

cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, glycogen storage disease, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, pseudohypoaldosteronism type 1, Liddle's syndrome, cystinuria, iminoglycinuria, Hartup disease, Fanconi disease, and Bartter syndrome; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, hemiplegic migraine, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, acid maltase deficiency (AMD, also known as Pompe's disease), generalized myotonia, and myotonia congenita; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-

candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding TRICH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRICH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or

amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRICH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis

methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of

transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present

invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such

cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state

with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH.

In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/243,989, U.S. Ser. No. 60/245,904, U.S. Ser. No. 60/249,661, U.S. Ser. No. 60/247,673, U.S. Ser. No. 60/252,232, and U.S. Ser. No. 60/250,790, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the

ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and

threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:21-40. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative transporters and ion channels were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly f Gen mic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of TRICH Encoding P lynucle tides

The sequences which were used to assemble SEQ ID NO:21-40 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other

implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:21-40 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is

calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of TRICH Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing

primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:21-40 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate

filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

XII. Expression f TRICH

Expression and purification of TRICH is achieved using bacterial or virus-based expression

systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

TRICH function is assessed by expressing the sequences encoding TRICH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome

formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TRICH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of TRICH Specific Antibodies

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH activity by, for example, binding the peptide or TRICH to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring TRICH Using Specific Antibodies

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRICH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

XVI. Identification of Molecules Which Interact with TRICH

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as Gβγ proteins (Reimann, supra) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, supra). TRICH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) Nature 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or lexA, and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the TRICH interacting proteins (fusion proteins with an activation domain) reconstitute a transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, Meth. Enzymol. 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions

between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH ion channel activity using the assays described in section XVIII.

XVII. Demonstration of TRICH Activity

Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as ß-galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and ß-galactosidase.

Transformed cells expressing ß-galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or ß-galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel, and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing Xenopus laevis oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., supra; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate Xenopus oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18 °C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of the TRICH mediator, such as cAMP, cGMP, or Ca^{+2} (in the form of $CaCl_2$), where appropriate. Electrode resistance is set at 2-5 MΩ and electrodes are filled with the intracellular solution lacking mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to the activity of TRICH in the assay.

In particular, the activity of TRICH-2 is measured as voltage-gated Ca^{2+} or Na^+ conductance, the activity of TRICH-15 is measured as Ca^{2+} conductance, and the activity of TRICH-16 is measured as K^+ conductance.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates into Xenopus laevis oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 1mM Na_2HPO_4, 5 mM Hepes, 3.8 mM NaOH , 50μg/ml gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g. radiolabeled with ^3H, fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+-free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate. In particular, test substrates include tricarboxylates for TRICH-1, H^+ for TRICH-3, sulfate for TRICH-4, Na^+ for TRICH-5, anionic metabolites for TRICH-6, glucose-6-phosphate for TRICH-8, and amino acids for TRICH-10.

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP-[γ-^{32}P], separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered ^{32}P using a scintillation counter. The reaction mixture contains ATP-[γ-^{32}P] and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of ^{32}P liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

Lipocalin activity of TRICH is measured by ligand fluorescence enhancement spectrofluorometry (Lin et al. (1997) Molecular Vision 3:17). Examples of ligands include retinol (Sigma, St. Louis MO) and 16-anthryloxy-palmitic acid (16-AP) (Molecular Probes Inc., Eugene OR). Ligand is dissolved in 100% ethanol and its concentration is estimated using known extinction coefficents (retinol: 46,000 A/M/cm at 325 nm; 16-AP: 8,200 A/M/cm at 361 nm). A 700 μl aliquot of 1 μM TRICH in 10 mM Tris (pH 7.5), 2 mM EDTA, and 500 mM NaCl is placed in a 1 cm path length quartz cuvette and 1 μl aliquots of ligand solution are added. Fluorescence is measured 100 seconds after each addition until readings are stable. Change in fluorescence per unit change in ligand concentration is proportional to TRICH activity.

XVIII. Identificati n f TRICH Agonists and Antag nists

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK

(Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVII. Alternatively, ion channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) Meth. Enzymol. 294:20-47; West, M.R. and C.R. Molloy (1996) Anal. Biochem. 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal ion concentration are measured using fluorescent dyes such as the Ca^{2+} indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the Cl^- indicator MQAE (all available from Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonyl dyes such as $DiBAC_4$ (Molecular Probes). $DiBAC_4$ equilibrates between the extracellular solution and cellular sites according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of $DiBAC_4$ entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) Curr. Opin. Biotechnol. 9:624-631). Candidate agonists or antagonists may be selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1626101	1	1626101CD1	21	1626101CB1
2907828	2	2907828CD1	22	2907828CB1
3968527	3	3968527CD1	23	3968527CB1
7472732	4	7472732CD1	24	7472732CB1
7476938	5	7476938CD1	25	7476938CB1
8128531	6	8128531CD1	26	8128531CB1
7476757	7	7476757CD1	27	7476757CB1
266243	8	266243CD1	28	266243CB1
6585710	9	6585710CD1	29	6585710CB1
7483599	10	7483599CD1	30	7483599CB1
2507246	11	2507246CD1	31	2507246CB1
3033505	12	3033505CD1	32	3033505CB1
4027693	13	4027693CD1	33	4027693CB1
7472030	14	7472030CD1	34	7472030CB1
7476089	15	7476089CD1	35	7476089CB1
6428177	16	6428177CD1	36	6428177CB1
7477243	17	7477243CD1	37	7477243CB1
7473042	18	7473042CD1	38	7473042CB1
7482060	19	7482060CD1	39	7482060CB1
1578772	20	1578772CD1	40	1578772CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	1626101CD1	g13785618	1.00E-105	[fl][Mus musculus] sideroflexin 4 Fleming, M. D. et al. (2001) A mutation in a mitochondrial transmembrane protein is responsible for the pleiotropic hematological and skeletal phenotype of flexed-tail (f/f) mice. Genes Dev. 15:652-657
		g545998	1.00E-15	[Rattus sp.] tricarboxylate carrier Azzi, A. et al. (1993) The mitochondrial tricarboxylate carrier. J. Bioenerg. Biomembr. 25:515-524
2	2907828CD1	g12004581	0.0	[fl][Mus musculus] calcium channel
3	3968527CD1	g6434968	0.0	[Mus musculus] putative E1-E2 ATPase Halleck, M.S. et al. (1999) Differential expression of putative transbilayer amphipath transporters. Physiol. Genomics (Online) 1:139-150
4	7472732CD1	g15341552	0.0	[fl][Homo sapiens] (AF331521) putative anion transporter
		g575895	1.40E-82	[Mus musculus] sulfate transporter Kobayashi, T. et al. (1997) Cloning of mouse diastrophic dysplasia sulfate transporter gene induc d during osteoblast differentiation by bone morphogenetic protein-2. Gene 198:341-349
5	7476938CD1	g4894991	6.00E-139	[Drosophila melanogaster] sodium-hydrogen exchanger NHE1
6	8128531CD1	g5052319	6.00E-51	[Homo sapiens] citrin; adult-onset type II citrullinemia protein Kobayashi, K. et al. (1999) The gene mutated in adult-onset type II citrullinaemia encodes a putative mitochondrial carrier protein. Nat. Genet. 22:159-163
7	7476757CD1	g1653342	1.80E-16	[Synechocystis sp.] melibiose carrier protein Kaneko, T. et al. (1996) DNA Res. 3:109-136
8	266243CD1	g7229675	6.90E-39	[Arabidopsis thaliana] glucose 6 phosphate/phosphate translocator

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
9	6585710CD1	g15146444	0.0	[fl][Homo sapiens] sterolin-2 Lu,K. et al. (2001) Two genes that map to the stsl locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by ABCG5 and ABCG8, respectively. Am. J. Hum. Genet. 69:278-290
11	2507246CD1	g472900	3.10E-66	[Caenorhabditis elegans] carrier protein (c2) Runswick,M.J. et al. (1994) Extension of the mitochondrial transport superfamily: sequences of five members from the nematode worm Caenorhabditis elegans. DNA Seq. 4:281-291
12	3033505CD1	g6978016	4.40E-239	[Rattus norvegicus] neuronal glutamine transporter Varoqui,H. et al. (2000) Cloning and functional identification of a neuronal glutamine transporter. J. Biol. Chem. 275:4049-4054
13	4027693CD1	g2198807	1.10E-53	[Gallus gallus] monocarboxylate transporter 3 Yoon H. et al. (1997) Biochem. Biophys. Res. Commun. 234:90-94; Yoon H. and Philp N. (1998) J. Exp. Eye Res. 67:417-424; Yoon H. et al. (1999) Genomics 60:366-370
14	7472030CD1	g387426	0.0	[Mus musculus] multidrug resistance protein Gros, P. et al. (1986) Cell 47:371-380
15	7476089CD1	g2826759	2.50E-11	[Caenorhabditis elegans] sodium-calcium exchanger
16	6428177CD1	g3880445	1.70E-14	[Caenorhabditis elegans] contains similarity to Pfam domain: PF02214 (K+ channel tetramerisation domain)
17	7477243CD1	g6457274	0.0	[Mus musculus] putative E1-E2 ATPase Halleck,M.S. et al. (1999) Physiol. Genomics (Online) 1:139-150
18	7473042CD1	g15211933	0.0	[fl][Homo sapiens] proton myo-inositol transporter Uldry,M. et al. (2001) Identification of a mammalian H(+)-myo-inositol symporter expressed predominantly in the brain. The EMBO Journal 20:4467-4477

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
19	7482060CD1	g6006493	8.80E-83	[Homo sapiens] cardiac potassium channel subunit (Kv6.2)
20	1578772CD1	g11933425	2.00E-05	[fl][Arabidopsis thaliana] sulfate transporter
		g11907976	3.00E-05	[fl][Solanum tuberosum] high affinity sulfate transporter type 1

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1626101CD1	337	S221 S317 S324 T158 T34 T71	N154	PROTEIN TRANSMEMBRANE CHROMOSOME PUTATIVE TRANSPORTER C17G6.15C TRANSPORT XV READING FRAME PD006986:A20-P264	BLAST_PRODOM
2	2907828CD1	816	S123 S264 S351 S359 S375 S395 S4 S54 S697 S703 S716 S745 S769 T14 T322 T382 T559 T602 T618 T639 T764 Y624	N599 N611 N616 N695	Transmembrane domains: L107-L124, I235-F254, C297-F320, G506-L523, M560-F577, L666-I686	HMMER
					Ion transport protein domain: L437-I686	HMMER_PFAM
					Sodium channel signature PR00170: Q227-F254, S296-D325	BLIMPS_PRINTS
3	3968527CD1	1047	S1038 S179 S346 S366 S417 S453 S491 S498 S499 S548 S559 S605 S624 S629 S835 S920 T143 T147 T207 T212 T240 T276 T377 T390 T397 T445 T528 T634 T649 T665 T687 T707 T763 T776 T981 Y611	N182 N285 N535	Transmembrane domains: V300-V319, I953-M980, V1004-S1023	HMMER
					E1-E2 ATPase domain: G146-I174, N256-E279	HMMER_PFAM
					E1-E2 ATPases phosphorylation site signature BL00154: G158-F175, I385-F403, D653-L693	BLIMPS_BLOCKS
					E1-E2 ATPases phosphorylation site: T371-D419	PROFILESCAN
					P-type cation-transporting atpase superfamily signature PR00119: L389-F403, A669-D679	BLIMPS_PRINTS
					ATPASE HYDROLASE TRANSMEMBRANE PHOSPHORYLATION ATPBINDING PROTEIN PROBABLE CALCIUMTRANSPORTING CALCIUM TRANSPORT PD004657:S817-T1009 PD149930:C757-F816 PD006317:K149-D245	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					do ATPASE; CALCIUM; TRANSPORTING; DM02405|Q10309|127-865:Y495-S881, S152-A484, S91-V270 DM02405|P40527|208-977:R201-S881, S91-A482 DM02405|Q09891|206-1107:E715-L851, L444-F703, N154-R327, L336-G414 DM02405|P39524|236-1049:V148-V869, Q92-D365, I926-S1038	BLAST_DOMO
					E1-E2 ATPase motif: D391-T397	MOTIFS
4	7472732CD1	671	S10 S138 S225 S311 S345 S352 S494 S556 S640 S658 T507 T595	N125 N131 N661	Transmembrane domains: L228-N248, L399-Y417, V470-P488	HMMER
					Sulfate transporter family domain: M162-F487	HMMER_PFAM
					STAS (Sulfate Transporter and Anti Sigma factor antagonist) domain: E508-A652	HMMER_PFAM
					Sulfate transporter proteins signature BL01130: A150-M201, S53-L106	BLIMPS_BLOCKS
					PROTEIN TRANSPORT SULFATE TRANSPORTER TRANSMEMBRANE PERMEASE INTERGENIC REGION AFFINITY GLYCOPROTEIN PD001255:V164-R486 PD001121:P33-G168	BLAST_PRODOM
					SULFATE TRANSPORTERS DM01229|P40879|5-462:P33-W446 DM01229|P50443|49-505:L32-N447 DM01229|P45380|10-468:R6-W446 DM01229|P53393|11-447:P33-M201, H303-W446	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7476938CD1	671	S15 S288 S582 S661 S663 S68 S85 T104 T120 T334 T440 T503 T584 T600 T630 T634 T649	N103 N110 N276 N337 N47 N580	Transmembrane domains: M150-R171, Y172-I194, I241-I263, G390-G418, V452-L480, L537-G555	HMMER
					Sodium/hydrogen exchanger family domain: I152-D568	HMMER_PFAM
					Na+/H+ exchanger signature PR01084: M215-F226, G229-S243, I244-T252, G284-A294	BLIMPS_PRINTS
					+ TRANSPORT EXCHANGER NA PD01672: M215-I263, A297-L333	BLIMPS_PRODOM
					NA+/H+ PROTEIN TRANSMEMBRANE TRANSPORT ANTIPORTER SYMPORT SODIUM EXCHANGER GLYCOPROTEIN SODIUM/HYDROGEN PD000631:G149-E567	BLAST_PRODOM
					do BETA; EXCHANGER; NA; DM02572|Q01345|12-703:L157-N580 DM02572|P48761|17-738:S155-N580 DM02572|P26434|14-716:L156-R632 DM02572|P48764|10-734:L156-E636	BLAST_DOMO
6	8128531CD1	315	S165 T149 T160 T2 T240 T251 T55 T9 Y261	N71	Mitochondrial carrier proteins domain: S7-Q99, N101-G217	HMMER_PFAM
					Mitochondrial energy transfer proteins signature BL00215: I13-Q37, L173-G185	BLIMPS_BLOCKS
					Mitochondrial energy transfer proteins signature: H3-T55, L102-S150, F221-L271	PROFILESCAN
					PROTEIN TRANSPORT TRANSMEMBRANE REPEAT MITOCHONDRION CARRIER MEMBRANE INNER MITOCHONDRIAL ADP/ATP PD000117:S7-Y302	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026|S54495|534-620:G15-G97 DM00026|S54495|622-719:E105-Q131, A170-L207 DM00026|S60949|16-113:I13-L92	BLAST_DOMO
					Mitochondrial carrier protein motif: P28-L36 P241-I249	MOTIFS
7	7476757CD1	445	S247 S26 S389 S393 S440 T181 T355 T380 T405		SODIUM:GALACTOSIDE SYMPORTER FAMILY DM01084|P30868|1-456:L178-K382 (p=6.1e-07)	BLAST_DOMO
					TRANSPORT PROTEIN TRANSMEMBRANE SYMPORT SUGAR SYMPORTER PERMEASE INNER MEMBRANE CARRIER PD003362:R183-L375 (p=5.7e-09)	BLAST_PRODOM
					signal peptide signal_peptide:M1-G29	HMMER
8	266243CD1	410	S162 S208 S356 S363 T140 T180 T297 T335 T351	N295 N333	Integral membrane protein DUF6 DUF6: A39-V181	HMMER_PFAM
					INTERMEMBRANE SPACE DOMAIN DM02684|P52178|1-401: V112-K336, P15-L63 DM02684|S37550|1-407: R110-K345, L36-V88 DM02684|S37497|1-409: P93-K345, H10-G90 DM02684|P52177|1-406: R110-K345, V6-S89	BLAST_DOMO
					TRANSLOCATOR PRECURSOR TRANSMEMBRANE CTPT PHOSPHATE/PHOSPHOENOLPYRUVATE PYRUVATE TRIOSE PHOSPHATE/PHOSPHATE NONGREEN PLASTID PD150555: L184-K336	BLAST_PRODOM
					transmembrane_domain: L169-F186, S316-Y332	HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databas s
9	6585710CD1	374	S145 S174 S310 S332 S362 T26 T313 T73 T80	N280 N320	do WHITE; FRUIT; FLY; SCARLET; DM05200|P45844|289-650: M1-V348 DM05200|P10090|317-666: V2-L346	BLAST_DOMO
					do PERMEASE; DEPENDENT; ATP; PDR10; DM01528|P51533|406-797: Y4-P340 DM01528|S55517|406-797: Y4-P340	BLAST_DOMO
					transmembrane_domain: L118-F137, Y339-V361	HMMER
10	7483599CD1	443	S412 S59 T26 T369 T388	N259	Transmembrane amino acid transporter Aa_trans: A104-F438	HMMER_PFAM
					ACID AMINO PROTEIN TRANSPORTER PERMEASE TRANSMEMBRANE INTERGENIC REGION PROLINE PD001875: G79-L367	BLAST_PRODOM
					transmembrane_domain: V115-C134, V177-F195, Y231-F254, F292-L310, L326-G344	HMMER
11	2507246CD1	321	S209 S307 S41 S80 T226 T274 Y268	N224 N229	signal_cleavage: M1-G25	SPSCAN
					Mitochondrial carrier proteins: N10-P125, S127-A220, S232-X322	HMMER_PFAM
					Mitochondrial energy transfer proteins BL00215:F16-Q40, V177-G189	BLIMPS_BLOCKS
					Mitochondrial energy transfer proteins signature: Q6-V101, Q6-G100, Q6-V99, Q6-L98, Q6-N97, Q6-P96, Q6-G95, Q6-L94, Q6-K84, F233-G283	PROFILESCAN
					TRANSPORT TRANSMEMBRANE MITOCHONDRION CARRIER INNER MITOCHONDRIAL ADP/ATP PD000117:L171-E317, H14-E216, S127-F292, S127-E264, L12-S232, R184-T320, N10-E117, T231-E310	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026|S44092|201-284:P125-L214, G73-I122, M237-E310, N10-Q40 DM00026|S44092|302-380:M237-L316, S134-E208, T68-Y112, L15-L82 DM00026|P38127|56-163:A17-F119, E263-Y309 DM00026|P38127|291-375:A239-L316, S134-L214, G73-Y112, A17-S42	BLAST_DOMO
					Mitochondrial energy transfer proteins signature: P31-L39 P253-L261	MOTIFS
12	3033505CD1	487	S303 S347 S378 S413 S45 S481 S482 S49 S56 S6 T145 T17 T259 T265 T32 T332 T355 T374 T442 T450	N15 N23 N251 N257 N26 N312 N79	Transmembrane domains: M85-L107, V198-T221, F224-I243, F316-N336, L352-F372, I399-V421, L457-W477	HMMER
					Transmembrane amino acid transporter: A95-S469	HMMER_PFAM
					ACID AMINO TRANSPORTER PERMEASE TRANSMEMBRANE INTERGENIC PROLINE PD001875:S76-V370	BLAST_PRODOM
					TRANSPORTER PROTEIN PD138374:H343-H487	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	4027693CD1	509	S233 T267 T419 Y277	N411	Transmembrane domain: P344-F366, L400-V418, F464-L483	HMMER
					Monocarboxylate transporter: S19-S467	HMMER_PFAM
					TRANSPORTER; LINKED DM05037|Q03064|1-475:P7-N244, V298-L480 DM05037|P53988|1-465:P7-S194, A299-L480 DM05037|P36021|155-612:P7-Y211, L300-L483	BLAST_DOMO
14	7472030CD1	1232	S1081 S1111 S1159 S1165 S253 S369 S401 S426 S554 S651 S654 S673 S887 T1178 T1186 T1211 T137 T20 T233 T491 T586 T595 T640 T728 T886 Y505 Y875	N1079 N1163 N189 N300 N372 N391 N424 N703 N764 N794 N86 N885 N92	Transmembrane domains: V112-I131, I735-Y751, M812-M834	HMMER
					ABC transporter transmembrane region: M49-I340, V693-I943	HMMER-PFAM
					ABC transporter: G1018-G1204, G415-G599	HMMER-PFAM
					ABC transporters family signatures: M526-L540, L1131-L1145	MOTIFS
					ATP/GTP-binding site motif A (P-loop): G422-S429, G1025-S1032	MOTIFS
					ABC transporters family signatures: I508-D557, I1113-D1162	ProfileScan
					ABC transporters family BL00211: L420-V431, L1131-D1162	BLIMPS-BLOCKS
					ATP-binding transport transmembrane protein PD00131: G162-D171, C1029-I1082, G1179-L1216	BLIMPS-PRODOM
					MALK protein DM00130|P21448|53-386: G52-G385, F748-G988, G696-N723	BLAST-DOMO
					MALK protein DM00130|S55692|70-399: L54-G385, F748-G988, T697-N723	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					MALK protein DM00130|P21439|61-391: L54-G385, F748-G988	BLAST-DOMO
					MALK protein DM00130|P23174|61-391: L54-G385, F748-G988, T697-N724	BLAST-DOMO
					ATP-binding transmembrane transporter, multidrug resistance, ABC transporter PD000130: L48-I334, F748-F933, V693-I718	BLAST-PRODOM
					P-glycoprotein, multidrug resistance, ATP-binding transporter PD167072: I465-A524	BLAST-PRODOM
					ATP-binding transmembrane transporter PD000101: E1058-G1128	BLAST-PRODOM
15	7476089CD1	759	S128 S161 S236 S315 S349 S405 S47 S505 S644 S646 S663 S710 S77 S84 T12 T130 T271 T435 T628 T735	N173 N184 N218 N250 N306 N334 N393 N461 N521 N545 N626 N682 N72 N729 N739	Transmembrane domain: I251-V267	HMMER
					Sodium/calcium exchanger: DM05297|P48765|6-969: V117-I346 (P-value = 6.0e-10)	BLAST-DOMO
16	6428177CD1	283	S163 S96 T119 Y112	N26 N87	K+ channel tetramerisation domain: V58-Q155	HMMER-PFAM
					Potassium channel signature: H100-T119	BLIMPS-PRINTS
					Potassium channel CDRK, SHAW: DM00490|P17972|1-102: V58-L143 (P-value = 2.8e-7)	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	7477243CD1	1129	S30 S47 S138 S278 S282 S442 S494 S495 S548 S708 S733 S736 S762 S813 S924 S982 S1094 S1100 S1105 S1109 S1113 T204 T250 T254 T264 T308 T328 T334 T408 T413 T449 T646 T680 T693 T701 T704 T1008 T1121 Y258 Y747	N121 N392 N761 N992 N1098	Transmembrane domains: F995-A1012, I1070-K1088	HMMER
					E1-E2 ATPases phosphorylation site proteins BL00154: G144-L161, V403-F421, K563-V573, D650-L690, T811-K834	BLIMPS_BLOCKS
					E1-E2 ATPases phosphorylation site: A389-V438	PROFILESCAN
					P-type cation-transporting atpase superfamily signature PR00119: F407-F421, A666-D676, I814-I833	BLIMPS_PRINTS
					ATPASE HYDROLASE TRANSMEMBRANE PHOSPHORYLA-TION ATP BINDING CALCIUM TRANSPORT PD004657: S848-K1088 PD149930: C787-Y847 PD006317: R135-I225	BLAST_PRODOM
					PROBABLE CALCIUM TRANSPORTING ATPASE HYDROLASE CALCIUM TRANSPORT TRANSMEM-BRANE PHOSPHORYLATION MAGNESIUM ATP BINDING PD101227: R458-V583, R16-L77	BLAST_PRODOM
					ATPASE; CALCIUM; TRANSPORTING DM02405|P39524|236-1049: Q79-L760, S708-N912, V989-F1024 DM02405|P32660|318-1225: E460-N912, E460-N912, I141-G446 DM02405|Q09891|206-1107: E460-N912, I141-G446, Y975-F1024 DM02405|S51243|356-1267: E460-Y911, E139-G446, V989-F1039, K453-G504	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ATP/GTP-binding site motif A (P-loop) A271-S278	MOTIFS
					E1-E2 ATPases phosphorylation site D409-T415	MOTIFS
18	7473042CD1	648	S2 S6 S224 S262 S294 S355 S419 S474 S607 S619 S640 S645 T12 T272 T287 T322 T394 T423 T461 Y11	N285 N433 N458 N485	Transmembrane domains: A79-F95, L366-V385, F395-S415	HMMER
					Sugar (and other) transporter: V84-F609	HMMER_PFAM
					Sugar transport proteins BL00216: G92-S103, L174-A223	BLIMPS_BLOCKS
					Sugar transport proteins signatures: L366-R421, S162-V225	PROFILESCAN
					Sugar transporter signature PR00171: G92-V102, L175-V194, Q336-Y346, L510-V531, S533-N545	BLIMPS_PRINTS
					Glucose transporter signature PR00172: L326-Y347, I364-V385,L90-K110, L510-S533, T543-L561, G574-L594	BLIMPS_PRINTS
					SUGAR TRANSPORTER PROTEIN PD000537: K296-R391	BLAST_PRODOM
					SUGAR TRANSPORT PROTEINS DM00135|Q01440|101-433: R178-S419, L513-G599 DM00135|P54723|120-454: R178-L414, T503-K600 DM00135|S25009|121-478: G161-Q417, L510-K598 DM00135|S43230|170-502: R178-Q417, S506-K600	BLAST_DOMO
					Sugar transport proteins signature 1 : G381-G396	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databas s
					Sugar transport proteins signature 2: V180-R205	MOTIFS
19	7482060CD1	545	S36 S43 S128 S198 S335 S405 T185 T269 T297 T441 Y95	N196	Transmembrane domains: I165-C182, V268-A285, V306-L325	HMMER
					Ion transport protein: I175-I390	HMMER_PFAM
					K+ channel tetramerisation domain: A9-L116	HMMER_PFAM
					Potassium channel signature PR00169: E60-G79, A157-T185, I205-K228, F231-V251, M275-C301, E304-E327, F339-M361, G368-F394	BLIMPS_PRINTS
					CHANNEL IONIC PROTEIN POTASSIUM SUBUNIT VOLTAGEGATED TRANSMEMBRANE CALCIUM TRANSPORT ION PD000141: F231-Y398	BLAST_PRODOM
					CHANNEL; POTASSIUM; CDRK; SHAW; DM00490|JH0595|26-142: V11-R115 DM00490|P15387|18-134: R5-R115 DM00436|JH0595|144-307: A163-I278 DM00490|P17970|268-384: V11-R115	BLAST_DOMO
20	1578772CD1	262	S87 S154 S232 S249 T153 T178		signal_cleavage: M1-G45	SPSCAN
					signal peptide: M1-A19	HMMER
					STAS domain (Sulphate Transporter and AntiSigma factor antagonist): H110-A236	HMMER_PFAM
					SULFATE TRANSPORTER PROTEIN TRANSPORT TRANSMEMBRANE AFFINITY GLYCOPROTEIN SULFATE HIGH DISEASE PD001755:H110-A236 E-value: 3.0e-08	BLAST_PRODOM

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
21	1626101CB1	1373	577-657	6122382H1 (BRAHNON05)	220	880
				2822668H1 (ADRETUT06)	1	305
				72008374V1	322	1036
				2378333T6 (ISLTNOT01)	822	1373
22	2907828CB1	3231	1-94, 2624-2681, 807-866	7276683H1 (LIVRDIS04)	2591	3231
				55084582J1	1325	1931
				55124156J1	574	1337
				7602868J1 (ESOGTME01)	468	1025
				7697278J1 (KIDPTDE01)	1914	2573
				7348609H1 (COLNNON05)	2022	2587
				7765470J1 (URETTUE01)	2378	3108
				6830849H1 (SINTNOR01)	1281	1833
				7376421H1 (ESOGTUE01)	1	582
23	3968527CB1	3160	2860-3160, 1-434, 1496-1790	GBI.g10277937_edit	1	429
				7068888H1 (BRATNOR01)	2443	3108
				7755687H1 (SPLNTUE01)	1389	2085
				7069701H1 (BRAUTDR02)	2679	3160
				7039903H1 (UTRSTMR02)	155	778
				7755687J1 (SPLNTUE01)	2047	2527
				8104892H1 (MIXDDIE02)	43	363
				55052339H1	380	1204
				7097441H1 (BRACDIR02)	2098	2643
				7032041H1 (BRAXTDR12)	1092	1580
24	7472732CB1	2848	2653-2848, 2455-2571, 817-1609, 1-72, 150-195	GBI.g8748893_000007.edit	1919	2722
				GNN.g6598919_006.edit	677	1384
				GBI.g8748893_000006_000003.edit	1271	1558
				g3179340	1705	2116
				GBI.g8748893_000003_000004.regenscan.edit	1463	1871

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5′ Position	3′ Position
				55061545H1	1	427
				55061546H1	116	728
				2467913T6 (THYRNOT08)	2098	2755
				2467913F6 (THYRNOT08)	1405	1945
				6489280R6 (MIXDUNB01)	2234	2848
25	7476938CB1	3727	1-1490	4107326H1 (BRSTTUT17)	3169	3445
				5958480H1 (BRATNOT05)	2598	3198
				8242492H1 (BONEUNR01)	1879	2518
				71063602V1	1741	2405
				6932813H1 (SINTTMR02)	667	1257
				7171144H1 (BRSTTMC01)	196	613
				7226459H1 (LUNGTMC01)	2496	3097
				8144835J1 (MIXDTME01)	424	840
				6799132H1 (COLENOR03)	3188	3727
				8190281H1 (BMARTXN03)	1329	1773
				7963983H1 (SPLNFEA02)	1201	1697
				GNN.g9187761_004.edit	1	559
26	8128531CB1	2571	1-925	3187659H1 (THYMNON04)	2225	2571
				70030270D1	909	1312
				4860138F6 (BRSTTUT22)	953	1509
				2232088T6 (PROSNOT16)	1644	2204
				4341662H1 (BRAUNOT02)	1	266
				6883871J1 (BRAHTDR03)	477	945
				2232088F6 (PROSNOT16)	1991	2478
				4001257T6 (HNT2AZS07)	1372	1963
				8128531H1 (SCOMDIC01)	150	850
27	7476757CB1	1660	1107-1660, 490-803	GNN:g7712233_000033_002	667	1660
				55136433H1	1	737

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
28	266243CB1	2743	1-155, 2720-2743, 2490-2605	7765596J1 (URETTUE01)	2067	2743
				7629030H1 (GBLADIE01)	655	1237
				71153607V1	1700	2171
				7279379H1 (BMARTXE01)	1160	1778
				6618283J1 (BRAUTDR03)	563	1096
				7629030J1 (GBLADIE01)	6	660
				GNN:g8575919_008	1	1233
29	6585710CB1	3239	1-899, 2163-2217, 2516-2653	72460988D1	1862	2503
				72458459D1	1332	2068
				71978812V1	2433	3230
				72463146D1	580	1190
				72461256D1	1	669
				71977010V1	2584	3231
				72462439D1	968	1611
				71875053V1	2740	3239
30	7483599CB1	1615	749-823, 114-353	g2077361	1261	1615
				ENST00000023927	826	1096
				FL7483599_g7708819_000010_g7293314_1_2-3	243	420
				FL7483599_g7708819_000010_g7293314_1_3-4	334	517
				FL7483599_g7708819_000010_g7293314_1_4-5	421	606
				FL7483599_g7708819_000010_g7293314_1_7	679	825
				FL7483599_g7708819_000010_g7293314_1_5-6	521	663
				FL7483599_g7708819_000010_g7293314_1_9	925	1332
				g2077387	1015	1419

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
				GNN.g7417485_000010_002	1	1332
				FL7483599_g7708819_000010_g7293314_1_1-2	115	333
31	2507246CB1	1245	922-1245	71424096V1	300	1021
				504936R6 (TMLR3DT02)	739	1245
				354532F1 (RATRNOT01)	619	1237
				72229434D1	1	481
32	3033505CB1	4169	793-1236, 2297-2327, 4074-4169, 3031-3429	6045025J1 (BRABDIR02)	3278	3674
				7359286H1 (BRAIFEE05)	656	1046
				6547566H2 (PROSUNT01)	55	758
				4104913F6 (BRSTTUT17)	3125	3616
				4529404H1 (LYMBTXT01)	1	256
				71059135V1	2442	3042
				487605R6 (HNT2AGT01)	1194	1700
				6859847H1 (BRAIFEN08)	1945	2572
				4324588H1 (TLYMUNT01)	557	810
				7103133H1 (BRAWTDR02)	877	1301
				7288253H1 (BRAIFER06)	3584	4169
				6913384J1 (PITUDIR01)	2564	3174
				3033505F6 (TLYMNOT05)	1354	1913
				71246947V1	1864	2523
33	4027693CB1	3440	2635-2665, 558-1546	2866257F6 (KIDNNOT20)	2069	2700
				7364929H1 (OVARDIC01)	2923	3440
				7617003H1 (KIDNTUE01)	2423	3031
				5501287H1 (BRABDIR01)	450	614
				g1485147	1	451
				70503458V1	1705	2332
				7648768J1 (STOMTDE01)	1193	1833
				70618525V1	1489	2083
				8107676H1 (MIXDDIE02)	582	1038

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
				7363817H1 (OVARDIC01)	733	1273
				2866257H1 (KIDNNOT20)	2068	2402
				g3147430	211	561
34	7472030CB1	3699	2778-3235, 2347-2550, 1-982, 3286-3699, 1712-2259	1242602R6 (LUNGNOT03)	1359	1710
				FL7472030_g10445386_g307181_1_7-8	2347	2651
				FL7472030_g10445386_g307181_1_8-9	2551	2792
				56004293H1 (FLP600128)	1815	2263
				8180328H1 (EYERNON01)	941	1562
				GBI_g4508130_g10445386_edit	1	3699
35	7476089CB1	2428	1-861, 1225-2428	FL7476089_g8656012_g5902966	1300	2346
				58016676J1	1702	2428
				56003593J1	1	583
				58007776J1	897	1407
				7289568R6 (BRAIFER06)	512	967
36	6428177CB1	2243	1052-1089, 1737-2243	1991187F6 (CORPNOT02)	1264	1808
				7924964H1 (COLNTUS02)	1	592
				8059028J1 (LIVRTUE01)	616	1137
				2132191H1 (OVARNOT03)	1995	2243
				6437511H1 (BRAENOT02)	1412	1968
				7393426R8 (BRABDIE02)	351	986
				3234007H2 (COLNUCT03)	1862	2112
				8219858J2 (SINTFER02)	996	1698
37	7477243CB1	3711	1-673, 2513-3058, 3676-3711	55120512J1	252	1056
				55120612J1	1325	2135
				56000471J1	1	498
				8228596J1 (BRAUTDR02)	2417	3044
				6989392F7 (BRAIFER05)	1911	2665

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
				2863115T6 (KIDNNOT20)	3051	3711
				55155912J1	1730	2139
				5885787F8 (LIVRNON08)	2646	3327
				6258661F6 (BMARTXT06)	856	1444
38	7473042CB1	2717	1-607, 1886-2006	56008775J1	966	1762
				72622070V1	532	1238
				GNN.g7008856_000017_002	24	776
				71797555V1	1628	2287
				8195488H2 (BRAINOR03)	336	1087
				72457143D1	2069	2717
				55141001H1	1	253
				55061745J1	282	393
39	7482060CB1	2235	827-859, 1642-2235, 1525-1551	GNN.g9454649_000007_00 0012	428	2065
				6770140R8 (BRAUNOR01)	1	871
				5923423H1 (BRAIFET02)	1932	2235
				6770140F8 (BRAUNOR01)	1416	1648
40	1578772CB1	2563	710-845, 1-138, 2543-2563	5402544H1 (BRAHNOT01)	549	807
				1619843T6 (BRAITUT13)	1940	2563
				70880785V1	769	1319
				6449038H1 (BRAINOC01)	1275	1715
				2083536H1 (UTRSNOT08)	1122	1380
				6536573H1 (OVARDIN02)	1	475
				2755946R6 (THP1AZS08)	1445	1833
				6553201H1 (BRAFNON02)	1531	2121
				7239234H1 (BRAINOY02)	238	689

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
21	1626101CB1	NOSEDIN01
22	2907828CB1	UCMCL5T01
23	3968527CB1	SPLNTUE01
24	7472732CB1	THYRNOT08
25	7476938CB1	LIVRNOT03
26	8128531CB1	THYMNOT08
27	7476757CB1	ESOGTUE01
28	266243CB1	BMARTXE01
29	6585710CB1	SINTNOT21
31	2507246CB1	LATRTUT02
32	3033505CB1	TLYMNOT05
33	4027693CB1	KIDNNOT20
34	7472030CB1	LUNGNOT03
35	7476089CB1	BRAIFER06
36	6428177CB1	BRABDIE02
37	7477243CB1	BMARTXT06
38	7473042CB1	UTRENOT10
39	7482060CB1	BRAUNOR01
40	1578772CB1	BRAITUT12

Table 6

Library	Vector	Library Description
BMARTXE01	pINCY	This 5′ biased random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Schering AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microM for 8 hours.
BMARTXT06	pINCY	Library was constructed using RNA isolated from an untreated SH-SY5Y cell line derived from bone marrow neuroblastoma tumor cells removed from a 4-year-old Caucasian female.
BRABDIE02	pINCY	This 5′ biased random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day, for 40 years).
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAITUT12	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 40-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated grade 4 gemistocytic astrocytoma.
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a

Table 6 (cont.)

Library	Vector	Library Description
		microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
ESOGTUE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from esophageal tumor tissue removed from a 61-year-old Caucasian male during a partial esophagectomy, proximal gastrectomy, pyloromyotomy, and regional lymph node excision. Pathology indicated an invasive grade 3 adenocarcinoma in the esophagus, extending distally to involve the gastroesophageal junction. The tumor extended through the muscularis to involve periesophageal and perigastric soft tissues. One perigastric and two periesophageal lymph nodes were positive for tumor. There were multiple perigastric and periesophageal tumor implants. The patient presented with deficiency anemia and myelodysplasia. Patient history included hyperlipidemia, and tobacco and alcohol abuse in remission. Previous surgeries included adenotonsillectomy, rhinoplasty, vasectomy, and hemorrhoidectomy. A previous bone marrow aspiration found the marrow to be hypercellular for age and had a cellularity-to-fat ratio of 95:5. The marrow was focally densely fibrotic. Granulocytic precursors were slightly increased with normal maturation. The estimate of blast cells was greater than 5%. Megakaryocytes were increased and appeared atypical in clusters. Storage cells and granulomata were absent. Patient medications included Epoetin, Danocrine, Berocca Plus tablets, Selenium, vitamin B6 phosphate, vitamins E & C, and beta carotene. Family history included alcohol abuse, atherosclerotic coronary artery disease, type II diabetes, chronic liver disease, and primary cardiomyopathy in the father; and benign hypertension and cerebrovascular disease in the mother.
KIDNNOT20	pINCY	Library was constructed using RNA isolated from left kidney tissue removed from a 43-year-old Caucasian male during nephroureterectomy, regional lymph node excision, and unilateral left adrenalectomy. Pathology for the associated tumor tissue indicated a grade 2 renal cell carcinoma. Family history included atherosclerotic coronary artery disease.

Table 6 (cont.)

Library	Vector	Library Description
LATRTUT02	pINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
LIVRNOT03	pINCY	Library was constructed using RNA isolated from liver tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20 weeks' gestation.
LUNGNOT03	PSPORT1	Library was constructed using RNA isolated from lung tissue of a 79-year-old Caucasian male. Pathology for the associated tumor tissue indicated grade 4 carcinoma. Patient history included a benign prostate neoplasm and atherosclerosis.
NOSEDIN01	pINCY	This normalized nasal polyp tissue library was constructed from 1.08 million independent clones from a pooled nasal polyp tissue library. Starting RNA was made from pooled cDNA from two donors. cDNA was generated using mRNA isolated from a nasal polyp removed from a 78-year-old Caucasian male during nasal polypectomy (donor A) and from nasal polyps from another donor (donor B). Pathology (A) indicated a nasal polyp and striking eosinophilia, especially deep in the epithelium. In many instances, eosinophils were undergoing frank necrosis with striking deposition of Charcot-Leyden crystals. Foci of eosinophil infiltration in small islands of cells were seen in certain areas, and those areas closer to the appearance surface were losing definition and evidently undergoing necrosis. Examination of respiratory epithelium showed loss of surface epithelium in many areas, and there was a tendency for cells to aggregate around the epithelium. This nasal polyp showed typical histology for polypoid change associated with allergic disease. Patient history included asthma, allergy tests (which were positive for histamine but negative for common substances), a pulmonary function test (PFT, which showed reduction in the forced expiratory volume (FEV), with increase after use of a bronchodilator), and nasal polyps. Patient history (A) included asthma. Previous surgery (A) included a nasal polypectomy. The patient was not using glucocorticoids in treatment for asthma. The library was normalized in 1 round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo

Table 6 (cont.)

Library	Vector	Library Description
		et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
SINTNOT21	pINCY	Library was constructed using RNA isolated from small intestine tissue obtained from a 8-year-old Black male, who died from anoxia. Serology was negative.
SPLNTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from spleen tumor tissue removed from a 28-year-old male during total splenectomy. Pathology indicated malignant lymphoma, diffuse large cell type, B-cell phenotype with abundant reactive T-cells and marked granulomatous response involving the spleen, where it formed approximately 45 nodules, liver, and multiple lymph nodes.
THYMNOT08	pINCY	Library was constructed using RNA isolated from thymus tissue removed from a 4-month-old Caucasian male during a total thymectomy and open heart repair of atrioventricular canal defect using hypothermia. Pathology indicated a grossly normal thymus. The patient presented with a congenital heart anomaly, congestive heart failure, and Down's syndrome. Patient history included abnormal thyroid function study and premature birth. Previous procedures included right and left heart angiocardiography. Patient medications included Digoxin, Synthroid, and Lasix.
THYRNOT08	pINCY	Library was constructed using RNA isolated from the diseased left thyroid tissue removed from a 13-year-old Caucasian female during a complete thyroidectomy. Pathology indicated lymphocytic thyroiditis. Pathology for the matched tumor tissue indicated grade 1 papillary carcinoma. Multiple lymph nodes from the right, left, and midline section of the neck were negative for tumor. Fragments of the thymus were benign. Fibroadipose tissue was identified in the right inferior and superior parathyroid regions. Multiple lymph nodes (2 of 6) from the right side of the neck contained microscopic foci of metastatic papillary carcinoma. Patient history included attention deficit disorder with hyperactivity. Previous surgeries included an operative procedure on the external ear. Patient medications included Prozac. Family history included chronic obstructive asthma in the mother; alcohol abuse, benign hypertension, and depressive disorder in the grandparent(s); and attention deficit disorder with hyperactivity in the sibling(s).

Table 6 (cont.)

Library	Vector	Library Description
TLYMNOT05	pINCY	Library was constructed using RNA isolated from nonactivated Th2 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-4 in the presence of anti-IL-12 antibodies and B7-transfected COS cells.
UCMCL5T01	PBLUESCRIPT	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.
UTRENOT10	pINCY	Library was constructed using RNA isolated from pooled uterine endometiral tissue removed from three adult females during endometrial biopsy. Pathology indicated normal endometrium. All three patients were positive for Beta-3 integrin.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

19. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional

TRICH, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target

polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of TRICH in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant

immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

<110> INCYTE GENOMICS, INC.
TANG, Y. Tom
YUE, Henry
NGUYEN, Danniel B.
HAFALIA, April J.A.
ELLIOTT, Vicki S.
LU, Yan
WALIA, Narinder K.
YAO, Monique G.
BAUGHN, Mariah R.
GANDHI, Ameena R.
DING, Li
SANJANWALA, Madhusudan
RAMKUMAR, Jayalaxmi
ARVIZU, Chandra
GIETZEN, Kimberly J.
LAL, Preeti G.
AZIMZAI, Yalda
KHAN, Farrah A.
THANGAVELU, Kavitha
THORNTON, Michael
LU, Dyung Aina M.
TRIBOULEY, Catherine M.
WARREN, Bridget A.
ISON, H. Craig
DAS, Debopriya
RAUMANN, Brigette E.
POLICKY, Jennifer L.
KEARNEY, Liam

<120> TRANSPORTERS AND ION CHANNELS

<130> PI-0270 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/243,989; 60/245,904; 60/247,673; 60/249,661; 60/252,232
60/250,790
<151> 2000-10-27; 2000-11-03; 2000-11-09; 2000-11-17; 2000-11-20;
2000-12-01

<160> 40

<170> PERL Program

<210> 1
<211> 337
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1626101CD1

<400> 1
Met Ser Leu Glu Gln Glu Glu Glu Thr Gln Pro Gly Arg Leu Leu
1 5 10 15
Gly Arg Arg Asp Ala Val Pro Ala Phe Ile Glu Pro Asn Val Arg
20 25 30
Phe Trp Ile Thr Glu Arg Gln Ser Phe Ile Arg Arg Phe Leu Gln
35 40 45
Trp Thr Glu Leu Leu Asp Pro Thr Asn Val Phe Ile Ser Val Glu

```
                50                  55                  60
Ser Ile Glu Asn Ser Arg Gln Leu Leu Cys Thr Asn Glu Asp Val
                65                  70                  75
Ser Ser Pro Ala Ser Ala Asp Gln Arg Ile Gln Glu Ala Trp Lys
                80                  85                  90
Arg Ser Leu Ala Thr Val His Pro Asp Ser Ser Asn Leu Ile Pro
                95                 100                 105
Lys Leu Phe Arg Pro Ala Ala Phe Leu Pro Phe Met Ala Pro Thr
               110                 115                 120
Val Phe Leu Ser Met Thr Pro Leu Lys Gly Ile Lys Ser Val Ile
               125                 130                 135
Leu Pro Gln Val Phe Leu Cys Ala Tyr Met Ala Ala Phe Asn Ser
               140                 145                 150
Ile Asn Gly Asn Arg Ser Tyr Thr Cys Lys Pro Leu Glu Arg Ser
               155                 160                 165
Leu Leu Met Ala Gly Ala Val Ala Ser Ser Thr Phe Leu Gly Val
               170                 175                 180
Ile Pro Gln Phe Val Gln Met Lys Tyr Gly Leu Thr Gly Pro Trp
               185                 190                 195
Ile Lys Arg Leu Leu Pro Val Ile Phe Leu Val Gln Ala Ser Gly
               200                 205                 210
Met Asn Val Tyr Met Ser Arg Ser Leu Glu Ser Ile Lys Gly Ile
               215                 220                 225
Ala Val Met Asp Lys Glu Gly Asn Val Leu Gly His Ser Arg Ile
               230                 235                 240
Ala Gly Thr Lys Ala Val Arg Glu Thr Leu Ala Ser Arg Ile Val
               245                 250                 255
Leu Phe Gly Thr Ser Ala Leu Ile Pro Glu Val Phe Thr Tyr Phe
               260                 265                 270
Phe Lys Arg Thr Gln Tyr Phe Arg Lys Asn Pro Gly Ser Leu Trp
               275                 280                 285
Ile Leu Lys Leu Ser Cys Thr Val Leu Ala Met Gly Leu Met Val
               290                 295                 300
Pro Phe Ser Phe Ser Ile Phe Pro Gln Ile Gly Gln Ile Gln Tyr
               305                 310                 315
Cys Ser Leu Glu Glu Lys Ile Gln Ser Pro Thr Glu Glu Thr Glu
               320                 325                 330
Ile Phe Tyr His Arg Gly Val
               335

<210> 2
<211> 816
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2907828CD1

<400> 2
Met Ala Val Ser Leu Asp Asp Asp Val Pro Leu Ile Leu Thr Leu
  1                 5                  10                  15
Asp Glu Gly Gly Ser Ala Pro Leu Ala Pro Ser Asn Gly Leu Gly
                20                  25                  30
Gln Glu Glu Leu Pro Ser Lys Asn Gly Gly Ser Tyr Ala Ile His
                35                  40                  45
Asp Ser Gln Ala Pro Ser Leu Ser Ser Gly Gly Glu Ser Ser Pro
                50                  55                  60
Ser Ser Pro Ala His Asn Trp Glu Met Asn Tyr Gln Glu Ala Ala
                65                  70                  75
Ile Tyr Leu Gln Glu Gly Glu Asn Asn Asp Lys Phe Phe Thr His
                80                  85                  90
Pro Lys Asp Ala Lys Ala Leu Ala Ala Tyr Leu Phe Ala His Asn
```

```
                95                  100                 105
His Leu Phe Tyr Leu Met Glu Leu Ala Thr Ala Leu Leu Leu Leu
            110                 115                 120
Leu Leu Ser Leu Cys Glu Ala Pro Ala Val Pro Ala Leu Arg Leu
        125                 130                 135
Gly Ile Tyr Val His Ala Thr Leu Glu Leu Phe Ala Leu Met Val
    140                 145                 150
Val Val Phe Glu Leu Cys Met Lys Leu Arg Trp Leu Gly Leu His
155                 160                 165
Thr Phe Ile Arg His Lys Arg Thr Met Val Lys Thr Ser Val Leu
            170                 175                 180
Val Val Gln Phe Val Glu Ala Ile Val Val Leu Val Arg Gln Met
        185                 190                 195
Ser His Val Arg Val Thr Arg Ala Leu Arg Cys Ile Phe Leu Val
    200                 205                 210
Asp Cys Arg Tyr Cys Gly Gly Val Arg Arg Asn Leu Arg Gln Ile
215                 220                 225
Phe Gln Ser Leu Pro Pro Phe Met Asp Ile Leu Leu Leu Leu Leu
            230                 235                 240
Phe Phe Met Ile Ile Phe Ala Ile Leu Gly Phe Tyr Leu Phe Ser
        245                 250                 255
Pro Asn Pro Ser Asp Pro Tyr Phe Ser Thr Leu Glu Asn Ser Ile
    260                 265                 270
Val Ser Leu Phe Val Leu Leu Thr Thr Ala Asn Phe Pro Asp Val
275                 280                 285
Met Met Pro Ser Tyr Ser Arg Asn Pro Trp Ser Cys Val Phe Phe
            290                 295                 300
Ile Val Tyr Leu Ser Ile Glu Leu Tyr Phe Ile Met Asn Leu Leu
        305                 310                 315
Leu Ala Val Val Phe Asp Thr Phe Asn Asp Ile Glu Lys Arg Lys
    320                 325                 330
Phe Lys Ser Leu Leu Leu His Lys Arg Thr Ala Ile Gln His Ala
335                 340                 345
Tyr Arg Leu Leu Ile Ser Gln Arg Arg Pro Ala Gly Ile Ser Tyr
            350                 355                 360
Arg Gln Phe Glu Gly Leu Met Arg Phe Tyr Lys Pro Arg Met Ser
        365                 370                 375
Ala Arg Glu Arg Tyr Leu Thr Phe Lys Ala Leu Asn Gln Asn Asn
    380                 385                 390
Thr Pro Leu Leu Ser Leu Lys Asp Phe Tyr Asp Ile Tyr Glu Val
395                 400                 405
Ala Ala Leu Lys Trp Lys Ala Lys Lys Asn Arg Glu His Trp Phe
            410                 415                 420
Asp Glu Leu Pro Arg Thr Ala Leu Leu Ile Phe Lys Gly Ile Asn
        425                 430                 435
Ile Leu Val Lys Ser Lys Ala Phe Gln Tyr Phe Met Tyr Leu Val
    440                 445                 450
Val Ala Val Asn Gly Val Trp Ile Leu Val Glu Thr Phe Met Leu
455                 460                 465
Lys Gly Gly Asn Phe Phe Ser Lys His Val Pro Trp Ser Tyr Leu
            470                 475                 480
Val Phe Leu Thr Ile Tyr Gly Val Glu Leu Phe Leu Lys Val Ala
        485                 490                 495
Gly Leu Gly Pro Val Glu Tyr Leu Ser Ser Gly Trp Asn Leu Phe
    500                 505                 510
Asp Phe Ser Val Thr Val Phe Ala Phe Leu Gly Leu Leu Ala Leu
515                 520                 525
Ala Leu Asn Met Glu Pro Phe Tyr Phe Ile Val Val Leu Arg Pro
            530                 535                 540
Leu Gln Leu Leu Arg Leu Phe Lys Leu Lys Glu Arg Tyr Arg Asn
        545                 550                 555
Val Leu Asp Thr Met Phe Glu Leu Leu Pro Arg Met Ala Ser Leu
    560                 565                 570
```

```
Gly Leu Thr Leu Leu Ile Phe Tyr Tyr Ser Phe Ala Ile Val Gly
                575                 580                 585
Met Glu Phe Phe Cys Gly Ile Val Phe Pro Asn Cys Cys Asn Thr
                590                 595                 600
Ser Thr Val Ala Asp Ala Tyr Arg Trp Arg Asn His Thr Val Gly
                605                 610                 615
Asn Arg Thr Val Val Glu Glu Gly Tyr Tyr Tyr Leu Asn Asn Phe
                620                 625                 630
Asp Asn Ile Leu Asn Ser Phe Val Thr Leu Phe Glu Leu Thr Val
                635                 640                 645
Val Asn Asn Trp Tyr Ile Ile Met Glu Gly Val Thr Ser Gln Thr
                650                 655                 660
Ser His Trp Ser Arg Leu Tyr Phe Met Thr Phe Tyr Ile Val Thr
                665                 670                 675
Met Val Val Met Thr Ile Ile Val Ala Phe Ile Leu Glu Ala Phe
                680                 685                 690
Val Phe Arg Met Asn Tyr Ser Arg Lys Asn Gln Asp Ser Glu Val
                695                 700                 705
Asp Gly Gly Ile Thr Leu Glu Lys Glu Ile Ser Lys Glu Glu Leu
                710                 715                 720
Val Ala Val Leu Glu Leu Tyr Arg Glu Ala Arg Gly Ala Ser Ser
                725                 730                 735
Asp Val Thr Arg Leu Leu Glu Thr Leu Ser Gln Met Glu Arg Tyr
                740                 745                 750
Gln Gln His Ser Met Val Phe Leu Gly Arg Arg Ser Arg Thr Lys
                755                 760                 765
Ser Asp Leu Ser Leu Lys Met Tyr Gln Glu Glu Ile Gln Glu Trp
                770                 775                 780
Tyr Glu Glu His Ala Arg Glu Gln Glu Gln Gln Arg Gln Leu Ser
                785                 790                 795
Ser Ser Ala Ala Pro Ala Ala Gln Gln Pro Pro Gly Ser Arg Gln
                800                 805                 810
Arg Ser Gln Thr Val Thr
                815
```

<210> 3
<211> 1047
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3968527CD1

<400> 3

```
Met Thr Asp Asn Ile Pro Leu Gln Pro Val Arg Gln Lys Lys Arg
  1                 5                  10                  15
Met Asp Ser Arg Pro Arg Ala Gly Cys Cys Glu Trp Leu Arg Cys
                 20                  25                  30
Cys Gly Gly Gly Glu Ala Arg Pro Arg Thr Val Trp Leu Gly His
                 35                  40                  45
Pro Glu Lys Arg Asp Gln Arg Tyr Pro Arg Asn Val Ile Asn Asn
                 50                  55                  60
Gln Lys Tyr Asn Phe Phe Thr Phe Leu Pro Gly Val Leu Phe Asn
                 65                  70                  75
Gln Phe Lys Tyr Phe Phe Asn Leu Tyr Phe Leu Leu Leu Ala Cys
                 80                  85                  90
Ser Gln Phe Val Pro Glu Met Arg Leu Gly Ala Leu Tyr Thr Tyr
                 95                 100                 105
Trp Val Pro Leu Gly Phe Val Leu Ala Val Thr Val Ile Arg Glu
                110                 115                 120
Ala Val Glu Glu Ile Arg Cys Tyr Val Arg Asp Lys Glu Val Asn
                125                 130                 135
```

```
Ser Gln Val Tyr Ser Arg Leu Thr Ala Arg Gly Thr Val Lys Val
                140                 145                 150
Lys Ser Ser Asn Ile Gln Val Gly Asp Leu Ile Ile Val Glu Lys
            155                 160                 165
Asn Gln Arg Val Pro Ala Asp Met Ile Phe Leu Arg Thr Ser Glu
        170                 175                 180
Lys Asn Gly Ser Cys Phe Leu Arg Thr Asp Gln Leu Asp Gly Glu
    185                 190                 195
Thr Asp Trp Lys Leu Arg Leu Pro Val Ala Cys Thr Gln Arg Leu
200                 205                 210
Pro Thr Ala Ala Asp Leu Leu Gln Ile Arg Ser Tyr Val Tyr Ala
215                 220                 225
Glu Glu Pro Asn Ile Asp Ile His Asn Phe Val Gly Thr Phe Thr
230                 235                 240
Arg Glu Asp Ser Asp Pro Pro Ile Ser Glu Ser Leu Ser Ile Glu
245                 250                 255
Asn Thr Leu Trp Ala Gly Thr Val Val Ala Ser Gly Thr Val Val
260                 265                 270
Gly Val Val Leu Tyr Thr Gly Arg Glu Leu Arg Ser Val Met Asn
275                 280                 285
Thr Ser Asn Pro Arg Ser Lys Ile Gly Leu Phe Asp Leu Glu Val
290                 295                 300
Asn Cys Leu Thr Lys Ile Leu Phe Gly Ala Leu Val Val Val Ser
305                 310                 315
Leu Val Met Val Ala Leu Gln His Phe Ala Gly Arg Trp Tyr Leu
320                 325                 330
Gln Ile Ile Arg Phe Leu Leu Leu Phe Ser Asn Ile Ile Pro Ile
335                 340                 345
Ser Leu Arg Val Asn Leu Asp Met Gly Lys Ile Val Tyr Ser Trp
350                 355                 360
Val Ile Arg Arg Asp Ser Lys Ile Pro Gly Thr Val Val Arg Ser
365                 370                 375
Ser Thr Ile Pro Glu Gln Leu Gly Arg Ile Ser Tyr Leu Leu Thr
380                 385                 390
Asp Lys Thr Gly Thr Leu Thr Gln Asn Glu Met Ile Phe Lys Arg
395                 400                 405
Leu His Leu Gly Thr Val Ala Tyr Gly Leu Asp Ser Met Asp Glu
410                 415                 420
Val Gln Ser His Ile Phe Ser Ile Tyr Thr Gln Gln Ser Gln Asp
425                 430                 435
Pro Pro Ala Gln Lys Gly Pro Thr Leu Thr Thr Lys Val Arg Arg
440                 445                 450
Thr Met Ser Ser Arg Val His Glu Ala Val Lys Ala Ile Ala Leu
455                 460                 465
Cys His Asn Val Thr Pro Val Tyr Glu Ser Asn Gly Val Thr Asp
470                 475                 480
Gln Ala Glu Ala Glu Lys Gln Tyr Glu Asp Ser Cys Arg Val Tyr
485                 490                 495
Gln Ala Ser Ser Pro Asp Glu Val Ala Leu Val Gln Trp Thr Glu
500                 505                 510
Ser Val Gly Leu Thr Leu Val Gly Arg Asp Gln Ser Ser Met Gln
515                 520                 525
Leu Arg Thr Pro Gly Asp Gln Ile Leu Asn Phe Thr Ile Leu Gln
530                 535                 540
Ile Phe Pro Phe Thr Tyr Glu Ser Lys Arg Met Gly Ile Ile Val
545                 550                 555
Arg Asp Glu Ser Thr Gly Glu Ile Thr Phe Tyr Met Lys Gly Ala
560                 565                 570
Asp Val Val Met Ala Gly Ile Val Gln Tyr Asn Asp Trp Leu Glu
575                 580                 585
Glu Glu Cys Gly Asn Met Ala Arg Glu Gly Leu Arg Val Leu Val
590                 595                 600
Val Ala Lys Lys Ser Leu Ala Glu Glu Gln Tyr Gln Asp Phe Glu
```

```
                605                 610                 615
Ala Arg Tyr Val Gln Ala Lys Leu Ser Val His Asp Arg Ser Leu
                620                 625                 630
Lys Val Ala Thr Val Ile Glu Ser Leu Glu Met Glu Met Glu Leu
                635                 640                 645
Leu Cys Leu Thr Gly Val Glu Asp Gln Leu Gln Ala Asp Val Arg
                650                 655                 660
Pro Thr Leu Glu Thr Leu Arg Asn Ala Gly Ile Lys Val Trp Met
                665                 670                 675
Leu Thr Gly Asp Lys Leu Glu Thr Ala Thr Cys Thr Ala Lys Asn
                680                 685                 690
Ala His Leu Val Thr Arg Asn Gln Asp Ile His Val Phe Arg Leu
                695                 700                 705
Val Thr Asn Arg Gly Glu Ala His Leu Glu Leu Asn Ala Phe Arg
                710                 715                 720
Arg Lys His Asp Cys Ala Leu Val Ile Ser Gly Asp Ser Leu Glu
                725                 730                 735
Val Cys Leu Lys Tyr Tyr Glu Tyr Glu Phe Met Glu Leu Ala Cys
                740                 745                 750
Gln Cys Pro Ala Val Val Cys Cys Arg Cys Ala Pro Thr Gln Lys
                755                 760                 765
Ala Gln Ile Val Arg Leu Leu Gln Glu Arg Thr Gly Lys Leu Thr
                770                 775                 780
Cys Ala Val Gly Asp Gly Gly Asn Asp Val Ser Met Ile Gln Glu
                785                 790                 795
Ser Asp Cys Gly Val Gly Val Glu Gly Lys Glu Gly Lys Gln Ala
                800                 805                 810
Ser Leu Ala Ala Asp Phe Ser Ile Thr Gln Phe Lys His Leu Gly
                815                 820                 825
Arg Leu Leu Met Val His Gly Arg Asn Ser Tyr Lys Arg Ser Ala
                830                 835                 840
Ala Leu Ser Gln Phe Val Ile His Arg Ser Leu Cys Ile Ser Thr
                845                 850                 855
Met Gln Ala Val Phe Ser Ser Val Phe Tyr Phe Ala Ser Val Pro
                860                 865                 870
Leu Tyr Gln Gly Phe Leu Ile Ile Gly Tyr Ser Thr Ile Tyr Thr
                875                 880                 885
Met Phe Pro Val Phe Ser Leu Val Leu Asp Lys Asp Val Lys Ser
                890                 895                 900
Glu Val Ala Met Leu Tyr Pro Glu Leu Tyr Lys Asp Leu Leu Lys
                905                 910                 915
Gly Arg Pro Leu Ser Tyr Lys Thr Phe Leu Ile Trp Val Leu Ile
                920                 925                 930
Ser Ile Tyr Gln Gly Ser Thr Ile Met Tyr Gly Ala Leu Leu Leu
                935                 940                 945
Phe Glu Ser Glu Phe Val His Ile Val Ala Ile Ser Phe Thr Ser
                950                 955                 960
Leu Ile Leu Thr Glu Leu Leu Met Val Ala Leu Thr Ile Gln Thr
                965                 970                 975
Trp His Trp Leu Met Thr Val Ala Glu Leu Leu Ser Leu Ala Cys
                980                 985                 990
Tyr Ile Ala Ser Leu Val Phe Leu His Glu Phe Ile Asp Val Tyr
                995                1000                1005
Phe Ile Ala Thr Leu Ser Phe Leu Trp Lys Val Ser Val Ile Thr
               1010                1015                1020
Leu Val Ser Cys Leu Pro Leu Tyr Val Leu Lys Tyr Leu Arg Arg
               1025                1030                1035
Arg Phe Ser Pro Pro Ser Tyr Ser Lys Leu Thr Ser
               1040                1045
```

<210> 4
<211> 671
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472732CD1

<400> 4

```
Met Thr Gly Ala Lys Arg Lys Lys Lys Ser Met Leu Trp Ser Lys
  1               5                  10                  15
Met His Thr Pro Gln Cys Glu Asp Ile Ile Gln Trp Cys Arg Arg
             20                  25                  30
Arg Leu Pro Ile Leu Asp Trp Ala Pro His Tyr Asn Leu Lys Glu
             35                  40                  45
Asn Leu Leu Pro Asp Thr Val Ser Gly Ile Met Leu Ala Val Gln
             50                  55                  60
Gln Val Thr Gln Gly Leu Ala Phe Ala Val Leu Ser Ser Val His
             65                  70                  75
Pro Val Phe Gly Leu Tyr Gly Ser Leu Phe Pro Ala Ile Ile Tyr
             80                  85                  90
Ala Ile Phe Gly Met Gly His His Val Ala Thr Gly Thr Phe Ala
             95                 100                 105
Leu Thr Ser Leu Ile Ser Ala Asn Ala Val Glu Arg Ile Val Pro
            110                 115                 120
Gln Asn Met Gln Asn Leu Thr Thr Gln Ser Asn Thr Ser Val Leu
            125                 130                 135
Gly Leu Ser Asp Phe Glu Met Gln Arg Ile His Val Ala Ala Ala
            140                 145                 150
Val Ser Phe Leu Gly Gly Val Ile Gln Val Ala Met Phe Val Leu
            155                 160                 165
Gln Leu Gly Ser Ala Thr Phe Val Val Thr Glu Pro Val Ile Ser
            170                 175                 180
Ala Met Thr Thr Gly Ala Ala Thr His Val Val Thr Ser Gln Val
            185                 190                 195
Lys Tyr Leu Leu Gly Met Lys Met Pro Tyr Ile Ser Gly Pro Leu
            200                 205                 210
Gly Phe Phe Tyr Ile Tyr Ala Tyr Val Phe Glu Asn Ile Lys Ser
            215                 220                 225
Val Arg Leu Glu Ala Leu Leu Leu Ser Leu Leu Ser Ile Val Val
            230                 235                 240
Leu Val Leu Val Lys Glu Leu Asn Glu Gln Phe Lys Arg Lys Ile
            245                 250                 255
Lys Val Val Leu Pro Val Asp Leu Val Leu Ala Pro Asn Thr Ser
            260                 265                 270
Pro Leu His His His Tyr Asp Cys Leu Phe Ala Asn Phe Leu Glu
            275                 280                 285
Pro Pro Trp Glu Asp Gly Leu Pro Glu Gly Ala Phe Asn Gln Ala
            290                 295                 300
Glu Gly His Leu Arg Arg Asn Ile Ile Pro Ser Pro Arg Ala Pro
            305                 310                 315
Pro Met Asn Ile Leu Ser Ala Val Ile Thr Glu Ala Phe Gly Val
            320                 325                 330
Ala Leu Val Gly Tyr Val Ala Ser Leu Ala Leu Ala Gln Gly Ser
            335                 340                 345
Ala Lys Lys Phe Lys Tyr Ser Ile Asp Asp Asn Gln Glu Phe Leu
            350                 355                 360
Ala His Gly Leu Ser Asn Ile Val Ser Ser Phe Phe Phe Cys Ile
            365                 370                 375
Pro Ser Ala Ala Ala Met Gly Arg Thr Ala Gly Leu Tyr Ser Thr
            380                 385                 390
Gly Ala Lys Thr Gln Val Ala Cys Leu Ile Ser Cys Ile Phe Val
            395                 400                 405
Leu Ile Val Ile Tyr Ala Ile Gly Pro Leu Leu Tyr Trp Leu Pro
            410                 415                 420
```

```
Met Cys Val Leu Ala Ser Ile Ile Val Val Gly Leu Lys Gly Met
                425                 430                 435
Leu Ile Gln Phe Arg Asp Leu Lys Lys Tyr Trp Asn Val Asp Lys
                440                 445                 450
Ile Asp Trp Gly Ile Trp Val Ser Thr Tyr Val Phe Thr Ile Cys
                455                 460                 465
Phe Ala Ala Asn Val Gly Leu Leu Phe Gly Val Val Cys Thr Ile
                470                 475                 480
Ala Ile Val Ile Gly Arg Phe Pro Arg Ala Met Thr Val Ser Ile
                485                 490                 495
Lys Asn Met Lys Glu Met Glu Phe Lys Val Lys Thr Glu Met Asp
                500                 505                 510
Ser Glu Thr Leu Gln Gln Val Lys Ile Ile Ser Ile Asn Asn Pro
                515                 520                 525
Leu Val Phe Leu Asn Ala Lys Lys Phe Tyr Thr Asp Leu Met Asn
                530                 535                 540
Met Ile Gln Lys Glu Asn Ala Cys Asn Gln Pro Leu Asp Asp Ile
                545                 550                 555
Ser Lys Cys Glu Gln Asn Thr Leu Leu Asn Ser Leu Ser Asn Gly
                560                 565                 570
Asn Cys Asn Glu Glu Ala Ser Gln Ser Cys Pro Asn Glu Lys Cys
                575                 580                 585
Tyr Leu Ile Leu Asp Cys Ser Gly Phe Thr Phe Phe Asp Tyr Ser
                590                 595                 600
Gly Val Ser Met Leu Val Glu Val Tyr Met Asp Cys Lys Gly Arg
                605                 610                 615
Ser Val Asp Val Leu Leu Ala His Cys Thr Ala Ser Leu Ile Lys
                620                 625                 630
Ala Met Thr Tyr Tyr Gly Asn Leu Asp Ser Glu Lys Pro Ile Phe
                635                 640                 645
Phe Glu Ser Val Ser Ala Ala Ile Ser His Ile His Ser Asn Lys
                650                 655                 660
Asn Leu Ser Lys Leu Ser Asp His Ser Glu Val
                665                 670
```

<210> 5
<211> 671
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476938CD1

<400> 5

```
Met Val Met Glu Ala Gly Glu Ser Lys Gly Ile Val Leu Ser Ser
  1               5                  10                  15
Gly Lys Gly Leu His Ala Ala Ser Phe Met Val Glu Gly Glu Asn
                20                  25                  30
Val Arg Glu Gly Ile Gly Ser Glu Met Gly Thr Cys Pro Lys Trp
                35                  40                  45
Thr Asn Val Ser His Cys Lys Met Gly Ile Met Pro Val Leu Val
                50                  55                  60
Lys Gly Phe Val Leu Ser Gly Ser Arg Lys Gln Lys Arg Val Leu
                65                  70                  75
Leu Ala Pro Arg Leu Arg Thr Arg Trp Ser Trp Lys Leu Arg Arg
                80                  85                  90
Met Gly Glu Lys Met Ala Glu Glu Glu Arg Phe Pro Asn Thr Thr
                95                 100                 105
His Glu Gly Phe Asn Val Thr Leu His Thr Thr Leu Val Val Thr
               110                 115                 120
Thr Lys Leu Val Leu Pro Thr Pro Gly Lys Pro Ile Leu Pro Val
               125                 130                 135
```

```
Gln Thr Gly Glu Gln Ala Gln Gln Glu Glu Gln Ser Ser Gly Met
                140                 145                 150
Thr Ile Ph  Phe Ser Leu Leu Val Leu Ala Ile Cys Ile Ile Leu
            155                 160                 165
Val His Leu Leu Ile Arg Tyr Arg Leu His Phe Leu Pro Glu Ser
        170                 175                 180
Val Ala Val Val Ser Leu Gly Ile Leu Met Gly Ala Val Ile Lys
    185                 190                 195
Ile Ile Glu Phe Lys Lys Leu Ala Asn Trp Lys Glu Glu Glu Met
200                 205                 210
Phe Arg Pro Asn Met Phe Phe Leu Leu Leu Leu Pro Pro Ile Ile
215                 220                 225
Phe Glu Ser Gly Tyr Ser Leu His Lys Gly Asn Phe Phe Gln Asn
230                 235                 240
Ile Gly Ser Ile Thr Leu Phe Ala Val Phe Gly Thr Ala Ile Ser
245                 250                 255
Ala Phe Val Val Gly Gly Gly Ile Tyr Phe Leu Gly Gln Ala Asp
260                 265                 270
Val Ile Ser Lys Leu Asn Met Thr Asp Ser Phe Ala Phe Gly Ser
275                 280                 285
Leu Ile Ser Ala Val Asp Pro Val Ala Thr Ile Ala Ile Phe Asn
290                 295                 300
Ala Leu His Val Asp Pro Val Leu Asn Met Leu Val Phe Gly Glu
305                 310                 315
Ser Ile Leu Asn Asp Ala Val Ser Ile Val Leu Thr Asn Thr Ala
320                 325                 330
Glu Gly Leu Thr Arg Lys Asn Met Ser Asp Val Ser Gly Trp Gln
335                 340                 345
Thr Phe Leu Gln Ala Leu Asp Tyr Phe Leu Lys Met Phe Phe Gly
350                 355                 360
Ser Ala Ala Leu Gly Thr Leu Thr Gly Leu Ile Ser Ala Leu Val
365                 370                 375
Leu Lys His Ile Asp Leu Arg Lys Thr Pro Ser Leu Glu Phe Gly
380                 385                 390
Met Met Ile Ile Phe Ala Tyr Leu Pro Tyr Gly Leu Ala Glu Gly
395                 400                 405
Ile Ser Leu Ser Gly Ile Met Ala Ile Leu Phe Ser Gly Ile Val
410                 415                 420
Met Ser His Tyr Thr His His Asn Leu Ser Pro Val Thr Gln Ile
425                 430                 435
Leu Met Gln Gln Thr Leu Arg Thr Val Ala Phe Leu Cys Glu Thr
440                 445                 450
Cys Val Phe Ala Phe Leu Gly Leu Ser Ile Phe Ser Phe Pro His
455                 460                 465
Lys Phe Glu Ile Ser Phe Val Ile Trp Cys Ile Val Leu Val Leu
470                 475                 480
Phe Gly Arg Ala Val Asn Ile Phe Pro Leu Ser Tyr Leu Leu Asn
485                 490                 495
Phe Phe Arg Asp His Lys Ile Thr Pro Lys Met Met Phe Ile Met
500                 505                 510
Trp Phe Ser Gly Leu Arg Gly Ala Ile Pro Tyr Ala Leu Ser Leu
515                 520                 525
His Leu Asp Leu Glu Pro Met Glu Lys Arg Gln Leu Ile Gly Thr
530                 535                 540
Thr Thr Ile Val Ile Val Leu Phe Thr Ile Leu Leu Leu Gly Gly
545                 550                 555
Ser Thr Met Pro Leu Ile Arg Leu Met Asp Ile Glu Asp Ala Lys
560                 565                 570
Ala His Arg Arg Asn Lys Lys Asp Val Asn Leu Ser Lys Thr Glu
575                 580                 585
Lys Met Gly Asn Thr Val Glu Ser Glu His Leu Ser Glu Leu Thr
590                 595                 600
Glu Glu Glu Tyr Glu Ala His Tyr Ile Arg Arg Gln Asp Leu Lys
```

```
                605                 610                 615
Gly Phe Val Trp Leu Asp Ala Lys Tyr Leu Asn Pro Phe Phe Thr
                620                 625                 630
Arg Arg Leu Thr Gln Glu Asp Leu His His Gly Arg Ile Gln Met
                635                 640                 645
Lys Thr Leu Thr Asn Lys Trp Tyr Glu Glu Val Arg Gln Gly Pro
                650                 655                 660
Ser Gly Ser Glu Asp Asp Glu Gln Glu Leu Leu
                665                 670
```

<210> 6
<211> 315
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8128531CD1

<400> 6

```
Met Thr His Gln Asp Leu Ser Ile Thr Ala Lys Leu Ile Asn Gly
  1               5                  10                  15
Gly Val Ala Gly Leu Val Gly Val Thr Cys Val Phe Pro Ile Asp
                20                  25                  30
Leu Ala Lys Thr Arg Leu Gln Asn Gln His Gly Lys Ala Met Tyr
                35                  40                  45
Lys Gly Met Ile Asp Cys Leu Met Lys Thr Ala Arg Ala Glu Gly
                50                  55                  60
Phe Phe Gly Met Tyr Arg Gly Ala Ala Val Asn Leu Thr Leu Val
                65                  70                  75
Thr Pro Glu Lys Ala Ile Lys Leu Ala Ala Asn Asp Phe Phe Arg
                80                  85                  90
Arg Leu Leu Met Glu Asp Gly Met Gln Arg Asn Leu Lys Met Glu
                95                 100                 105
Met Leu Ala Gly Cys Gly Ala Gly Met Cys Gln Val Val Val Thr
               110                 115                 120
Cys Pro Met Glu Met Leu Lys Ile Gln Leu Gln Asp Ala Gly Arg
               125                 130                 135
Leu Ala Val His His Gln Gly Ser Ala Ser Ala Pro Ser Thr Ser
               140                 145                 150
Arg Ser Tyr Thr Thr Gly Ser Ala Ser Thr His Arg Arg Pro Ser
               155                 160                 165
Ala Thr Leu Ile Ala Trp Glu Leu Leu Arg Thr Gln Gly Leu Ala
               170                 175                 180
Gly Leu Tyr Arg Gly Leu Gly Ala Thr Leu Leu Arg Asp Ile Pro
               185                 190                 195
Phe Ser Ile Ile Tyr Phe Pro Leu Phe Ala Asn Leu Asn Asn Leu
               200                 205                 210
Gly Phe Asn Glu Leu Ala Gly Lys Ala Ser Phe Ala His Ser Phe
               215                 220                 225
Val Ser Gly Cys Val Ala Gly Ser Ile Ala Ala Val Ala Val Thr
               230                 235                 240
Pro Leu Asp Val Leu Lys Thr Arg Ile Gln Thr Leu Lys Lys Gly
               245                 250                 255
Leu Gly Glu Asp Met Tyr Ser Gly Ile Thr Asp Cys Ala Arg Lys
               260                 265                 270
Leu Trp Ile Gln Glu Gly Pro Ser Ala Phe Met Lys Gly Ala Gly
               275                 280                 285
Cys Arg Ala Leu Val Ile Ala Pro Leu Phe Gly Ile Ala Gln Gly
               290                 295                 300
Val Tyr Phe Ile Gly Ile Gly Glu Arg Ile Leu Lys Cys Phe Asp
               305                 310                 315
```

<210> 7
<211> 445
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476757CD1

<400> 7
Met Pro Trp Val Leu Gly Cys Thr Pro Phe Ile Ala Leu Ala Tyr
1 5 10 15
Phe Phe Leu Trp Phe Leu Pro Pro Phe Thr Ser Leu Arg Gly Leu
20 25 30
Trp Tyr Thr Thr Phe Tyr Cys Leu Phe Gln Ala Leu Ala Thr Phe
35 40 45
Phe Gln Val Pro Tyr Thr Ala Leu Thr Met Leu Leu Thr Pro Cys
50 55 60
Pro Arg Glu Arg Asp Ser Ala Thr Ala Ile Pro Asp Asp Cys Gly
65 70 75
Asp Gly Gly Asn Thr Asp Gly Gly His Cys Pro Arg Ala His Arg
80 85 90
Val Arg Arg Pro Gln Thr Pro Gln Val Arg Gly His Cys Asp Pro
95 100 105
Gly Ala Ser His Cys Leu Pro Glu Cys Ser His Leu Tyr Cys Ile
110 115 120
Ala Ala Ala Val Val Val Val Thr Tyr Pro Val Cys Ile Ser Leu
125 130 135
Leu Cys Leu Gly Val Lys Glu Arg Pro Gly Phe Ala Phe Glu Leu
140 145 150
Cys Glu Ala Lys Val Thr Arg Phe Cys Val Ala Asp Pro Ser Ala
155 160 165
Pro Ala Ser Gly Pro Gly Leu Ser Phe Leu Ala Gly Leu Ser Leu
170 175 180
Thr Thr Arg His Pro Pro Tyr Leu Lys Leu Val Ile Ser Phe Leu
185 190 195
Phe Ile Ser Ala Ala Val Gln Val Glu Gln Ser Tyr Leu Val Leu
200 205 210
Phe Cys Thr His Ala Ser Gln Leu His Asp His Val Gln Gly Leu
215 220 225
Val Ser Ala Val Leu Ser Thr Pro Leu Trp Glu Trp Val Leu Gln
230 235 240
Arg Phe Gly Lys Lys Thr Ser Ala Phe Gly Ile Phe Ala Met Val
245 250 255
Pro Phe Ala Ile Leu Leu Ala Ala Val Pro Thr Ala Pro Val Ala
260 265 270
Tyr Val Val Ala Phe Val Ser Gly Val Ser Ile Ala Val Ser Leu
275 280 285
Leu Leu Pro Trp Ser Met Leu Pro Asp Val Val Asp Asp Phe Gln
290 295 300
Leu Gln His Arg His Gly Pro Gly Leu Glu Thr Ile Phe Tyr Ser
305 310 315
Ser Tyr Val Phe Phe Thr Lys Leu Ser Gly Ala Cys Ala Leu Gly
320 325 330
Ile Ser Thr Leu Ser Leu Glu Phe Ser Gly Tyr Lys Ala Gly Val
335 340 345
Cys Lys Gln Ala Glu Glu Val Val Val Thr Leu Lys Val Leu Ile
350 355 360
Gly Ala Val Pro Thr Cys Met Ile Leu Ala Gly Leu Cys Ile Leu
365 370 375
Met Val Gly Ser Thr Pro Lys Thr Pro Ser Arg Asp Ala Ser Ser
380 385 390
Arg Leu Ser Leu Arg Arg Arg Ala Gln Ala Pro Asn Val His Thr

```
                395                 400                 405
Ser Lys Val His Glu His Ala His Ile Met Gln Ala His Ala Gly
                410                 415                 420
Gln Ala Val Gly Gly Leu Val Ile Ser His Ser Leu Leu Arg Val
                425                 430                 435
Thr Ala Ser Gly Ser Ala Ala Glu Arg Tyr
                440                 445
```

<210> 8
<211> 410
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 266243CD1

<400> 8

```
Met Ala Ala Ala Ala Val Gly Ala Gly His Gly Ala Gly Gly Pro
  1               5                  10                  15
Gly Ala Ala Ser Ser Ser Gly Gly Ala Arg Glu Gly Ala Arg Val
                 20                  25                  30
Ala Ala Leu Cys Leu Leu Trp Tyr Ala Leu Ser Ala Gly Gly Asn
                 35                  40                  45
Val Val Asn Lys Val Ile Leu Ser Ala Phe Pro Phe Pro Val Thr
                 50                  55                  60
Val Ser Leu Cys His Ile Leu Ala Leu Cys Ala Gly Leu Pro Pro
                 65                  70                  75
Leu Leu Arg Ala Trp Arg Val Pro Pro Ala Pro Pro Val Ser Gly
                 80                  85                  90
Pro Gly Pro Ser Pro His Pro Ser Ser Gly Pro Leu Leu Pro Pro
                 95                 100                 105
Arg Phe Tyr Pro Arg Tyr Val Leu Pro Leu Ala Phe Gly Lys Tyr
                110                 115                 120
Phe Ala Ser Val Ser Ala His Val Ser Ile Trp Lys Val Pro Val
                125                 130                 135
Ser Tyr Ala His Thr Val Lys Ala Thr Met Pro Ile Trp Val Val
                140                 145                 150
Leu Leu Ser Arg Ile Ile Met Lys Glu Lys Gln Ser Thr Lys Val
                155                 160                 165
Tyr Leu Ser Leu Ile Pro Ile Ile Ser Gly Val Leu Leu Ala Thr
                170                 175                 180
Val Thr Glu Leu Ser Phe Asp Met Trp Gly Leu Val Ser Ala Leu
                185                 190                 195
Ala Ala Thr Leu Cys Phe Ser Leu Gln Asn Ile Phe Ser Lys Lys
                200                 205                 210
Val Leu Arg Asp Ser Arg Ile His His Leu Arg Leu Leu Asn Ile
                215                 220                 225
Leu Gly Cys His Ala Val Phe Phe Met Ile Pro Thr Trp Val Leu
                230                 235                 240
Val Asp Leu Ser Ala Phe Leu Val Ser Ser Asp Leu Thr Tyr Val
                245                 250                 255
Tyr Gln Trp Pro Trp Thr Leu Leu Leu Leu Ala Val Ser Gly Phe
                260                 265                 270
Cys Asn Phe Ala Gln Asn Val Ile Ala Phe Ser Ile Leu Asn Leu
                275                 280                 285
Val Ser Pro Leu Ser Tyr Ser Val Ala Asn Ala Thr Lys Arg Ile
                290                 295                 300
Met Val Ile Thr Val Ser Leu Ile Met Leu Arg Asn Pro Val Thr
                305                 310                 315
Ser Thr Asn Val Leu Gly Met Met Thr Ala Ile Leu Gly Val Phe
                320                 325                 330
Leu Tyr Asn Lys Thr Lys Tyr Asp Ala Asn Gln Gln Ala Arg Lys
```

```
                335                 340                 345
His Leu Leu Pro Val Thr Thr Ala Asp Leu Ser Ser Lys Glu Arg
                350                 355                 360
His Arg Ser Pro Leu Glu Lys Pro His Asn Gly Leu Leu Phe Pro
                365                 370                 375
Gln His Gly Asp Tyr Gln Tyr Gly Arg Asn Asn Ile Leu Thr Asp
                380                 385                 390
His Phe Gln Tyr Ser Arg Gln Ser Tyr Pro Asn Ser Tyr Ser Leu
                395                 400                 405
Asn Arg Tyr Asp Val
                410
```

<210> 9
<211> 374
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6585710CD1

<400> 9

```
Met Val His Tyr Phe Thr Ala Ile Gly Tyr Pro Cys Pro Arg Tyr
  1                 5                  10                  15
Ser Asn Pro Ala Asp Phe Tyr Val Asp Leu Thr Ser Ile Asp Arg
                 20                  25                  30
Arg Ser Arg Glu Gln Glu Leu Ala Thr Arg Glu Lys Ala Gln Ser
                 35                  40                  45
Leu Ala Ala Leu Phe Leu Glu Lys Val Arg Asp Leu Asp Asp Phe
                 50                  55                  60
Leu Trp Lys Ala Glu Thr Lys Asp Leu Asp Glu Asp Thr Cys Val
                 65                  70                  75
Glu Ser Ser Val Thr Pro Leu Asp Thr Asn Cys Leu Pro Ser Pro
                 80                  85                  90
Thr Lys Met Pro Gly Ala Val Gln Gln Phe Thr Thr Leu Ile Arg
                 95                 100                 105
Arg Gln Ile Ser Asn Asp Phe Arg Asp Leu Pro Thr Leu Leu Ile
                110                 115                 120
His Gly Ala Glu Ala Cys Leu Met Ser Met Thr Ile Gly Phe Leu
                125                 130                 135
Tyr Phe Gly His Gly Ser Ile Gln Leu Ser Phe Met Asp Thr Ala
                140                 145                 150
Ala Leu Leu Phe Met Ile Gly Ala Leu Ile Pro Phe Asn Val Ile
                155                 160                 165
Leu Asp Val Ile Ser Lys Cys Tyr Ser Glu Arg Ala Met Leu Tyr
                170                 175                 180
Tyr Glu Leu Glu Asp Gly Leu Tyr Thr Thr Gly Pro Tyr Phe Phe
                185                 190                 195
Ala Lys Ile Leu Gly Glu Leu Pro Glu His Cys Ala Tyr Ile Ile
                200                 205                 210
Ile Tyr Gly Met Pro Thr Tyr Trp Leu Ala Asn Leu Arg Pro Gly
                215                 220                 225
Leu Gln Pro Phe Leu Leu His Phe Leu Leu Val Trp Leu Val Val
                230                 235                 240
Phe Cys Cys Arg Ile Met Ala Leu Ala Ala Ala Ala Leu Leu Pro
                245                 250                 255
Thr Phe His Met Ala Ser Phe Phe Ser Asn Ala Leu Tyr Asn Ser
                260                 265                 270
Phe Tyr Leu Ala Gly Gly Phe Met Ile Asn Leu Ser Ser Leu Trp
                275                 280                 285
Thr Val Pro Ala Trp Ile Ser Lys Val Ser Phe Leu Arg Trp Cys
                290                 295                 300
Phe Glu Gly Leu Met Lys Ile Gln Phe Ser Arg Arg Thr Tyr Lys
```

```
                305                 310                 315
Met Pro Leu Gly Asn Leu Thr Ile Ala Val Ser Gly Asp Lys Ile
                320                 325                 330
Leu Ser Ala Met Glu Leu Asp Ser Tyr Pro Leu Tyr Ala Ile Tyr
                335                 340                 345
Leu Ile Val Ile Gly Leu Ser Gly Gly Phe Met Val Leu Tyr Tyr
                350                 355                 360
Val Ser Leu Arg Phe Ile Lys Gln Lys Pro Ser Gln Asp Trp
                365                 370

<210> 10
<211> 443
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483599CD1

<400> 10
Met Asp Lys Phe Leu Asp Thr Tyr Asn Leu Pro Arg Leu Asn Gln
  1               5                  10                  15
Glu Glu Ile Gln Asn Leu Lys Arg Pro Ile Thr Ser Asn Glu Ile
                20                  25                  30
Lys Ala Ile Ile Lys Ser Leu Gln Met Ser Leu Leu Gly Arg Asp
                35                  40                  45
Tyr Asn Ser Glu Leu Asn Ser Leu Asp Asn Gly Pro Gln Ser Pro
                50                  55                  60
Ser Glu Ser Ser Ser Ser Ile Thr Ser Glu Asn Val His Pro Ala
                65                  70                  75
Gly Glu Ala Gly Leu Ser Met Met Gln Thr Leu Ile His Leu Leu
                80                  85                  90
Lys Cys Asn Ile Gly Thr Gly Leu Leu Gly Leu Pro Leu Ala Ile
                95                 100                 105
Lys Asn Ala Gly Leu Leu Val Gly Pro Val Ser Leu Leu Ala Ile
               110                 115                 120
Gly Val Leu Thr Val His Cys Met Val Ile Leu Leu Asn Cys Ala
               125                 130                 135
Gln His Leu Ser Gln Pro Arg Leu Gln Lys Thr Phe Val Asn Tyr
               140                 145                 150
Gly Glu Ala Thr Met Tyr Gly Leu Glu Thr Cys Pro Asn Thr Trp
               155                 160                 165
Leu Arg Ala His Ala Val Trp Gly Arg Tyr Thr Val Ser Phe Leu
               170                 175                 180
Leu Val Ile Thr Gln Leu Gly Phe Cys Ser Val Tyr Phe Met Phe
               185                 190                 195
Met Ala Asp Asn Leu Gln Gln Met Val Glu Lys Ala His Val Thr
               200                 205                 210
Ser Asn Ile Cys Gln Pro Arg Glu Ile Leu Thr Leu Thr Pro Ile
               215                 220                 225
Leu Asp Ile Arg Phe Tyr Met Leu Ile Ile Leu Pro Phe Leu Ile
               230                 235                 240
Leu Leu Val Phe Ile Gln Asn Leu Lys Val Leu Ser Val Phe Ser
               245                 250                 255
Thr Leu Ala Asn Ile Thr Thr Leu Gly Ser Met Ala Leu Ile Phe
               260                 265                 270
Glu Tyr Ile Met Glu Gly Ile Pro Tyr Pro Ser Asn Leu Pro Leu
               275                 280                 285
Met Ala Asn Trp Lys Thr Phe Leu Leu Phe Phe Gly Thr Ala Ile
               290                 295                 300
Phe Thr Phe Glu Gly Val Gly Met Val Leu Pro Leu Lys Asn Gln
               305                 310                 315
Met Lys His Pro Gln Gln Phe Ser Phe Val Leu Tyr Leu Gly Met
```

```
                320                 325                 330
Ser Ile Val Ile Ile Leu Tyr Ile Leu Leu Gly Thr Leu Gly Tyr
                335                 340                 345
Met Lys Phe Gly Ser Asp Thr Gln Ala Ser Ile Thr Leu Asn Leu
                350                 355                 360
Pro Asn Cys Trp Tyr Val Leu Pro Thr Ser Gly Glu Ile Gly Arg
                365                 370                 375
Asp Thr Gly Thr Val Leu Val Val Ile Ala Glu Ser Thr Ala Lys
                380                 385                 390
Leu Ser His Glu Ala Gly Asn Pro Ser Leu Glu Val Thr Tyr Val
                395                 400                 405
Ser Pro Ala His Thr Ala Ser Val Lys Ala Ser His Met Ala Ala
                410                 415                 420
Pro His Ser Lys Gly Ala Gly Lys Cys Asn Ser Ala Met Cys Leu
                425                 430                 435
Glu Val Phe Gly Glu Gln His Lys
                440
```

<210> 11
<211> 321
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2507246CD1

<400> 11

```
Met Ala Thr Gly Gly Gln Gln Lys Glu Asn Thr Leu Leu His Leu
  1                 5                  10                  15
Phe Ala Gly Gly Cys Gly Gly Thr Val Gly Ala Ile Phe Thr Cys
                 20                  25                  30
Pro Leu Glu Val Ile Lys Thr Arg Leu Gln Ser Ser Arg Leu Ala
                 35                  40                  45
Leu Arg Thr Val Tyr Tyr Pro Gln Val His Leu Gly Thr Ile Ser
                 50                  55                  60
Gly Ala Gly Met Val Arg Pro Thr Ser Val Thr Pro Gly Leu Phe
                 65                  70                  75
Gln Val Leu Lys Ser Ile Leu Glu Lys Glu Gly Pro Lys Ser Leu
                 80                  85                  90
Phe Arg Gly Leu Gly Pro Asn Leu Val Gly Val Ala Pro Ser Arg
                 95                 100                 105
Ala Val Tyr Phe Ala Cys Tyr Ser Lys Ala Lys Glu Gln Phe Asn
                110                 115                 120
Gly Ile Phe Val Pro Asn Ser Asn Ile Val His Ile Phe Ser Ala
                125                 130                 135
Gly Ser Ala Ala Phe Ile Thr Asn Ser Leu Met Asn Pro Ile Trp
                140                 145                 150
Met Val Lys Thr Arg Met Gln Leu Glu Gln Lys Val Arg Gly Ser
                155                 160                 165
Lys Gln Met Asn Thr Leu Gln Cys Ala Arg Tyr Val Tyr Gln Thr
                170                 175                 180
Glu Gly Ile Arg Gly Phe Tyr Arg Gly Leu Thr Ala Ser Tyr Ala
                185                 190                 195
Gly Ile Ser Glu Thr Ile Ile Cys Phe Ala Ile Tyr Glu Ser Leu
                200                 205                 210
Lys Lys Tyr Leu Lys Glu Ala Pro Leu Ala Ser Ser Ala Asn Gly
                215                 220                 225
Thr Glu Lys Asn Ser Thr Ser Phe Phe Gly Leu Met Ala Ala Ala
                230                 235                 240
Ala Leu Ser Lys Gly Cys Ala Ser Cys Ile Ala Tyr Pro His Glu
                245                 250                 255
Val Ile Arg Thr Arg Leu Arg Glu Glu Gly Thr Lys Tyr Lys Ser
```

```
                260                 265                 270
Phe Val Gln Thr Ala Arg Leu Val Phe Arg Glu Glu Gly Tyr Leu
                275                 280                 285
Ala Phe Tyr Arg Gly Leu Phe Ala Gln Leu Ile Arg Gln Ile Pro
                290                 295                 300
Asn Thr Ala Ile Val Leu Ser Thr Tyr Glu Leu Ile Val Tyr Leu
                305                 310                 315
Leu Glu Asp Arg Thr Gln
                320
```

<210> 12
<211> 487
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3033505CD1

<400> 12

```
Met Met His Phe Lys Ser Gly Leu Glu Leu Thr Glu Leu Gln Asn
  1                 5                  10                  15
Met Thr Val Pro Glu Asp Asp Asn Ile Ser Asn Asp Ser Asn Asp
                 20                  25                  30
Phe Thr Glu Val Glu Asn Gly Gln Ile Asn Ser Lys Phe Ile Ser
                 35                  40                  45
Asp Arg Glu Ser Arg Arg Ser Leu Thr Asn Ser His Leu Glu Lys
                 50                  55                  60
Lys Lys Cys Asp Glu Tyr Ile Pro Gly Thr Thr Ser Leu Gly Met
                 65                  70                  75
Ser Val Phe Asn Leu Ser Asn Ala Ile Met Gly Ser Gly Ile Leu
                 80                  85                  90
Gly Leu Ala Phe Ala Leu Ala Asn Thr Gly Ile Leu Leu Phe Leu
                 95                 100                 105
Val Leu Leu Thr Ser Val Thr Leu Leu Ser Ile Tyr Ser Ile Asn
                110                 115                 120
Leu Leu Leu Ile Cys Ser Lys Glu Thr Gly Cys Met Val Tyr Glu
                125                 130                 135
Lys Leu Gly Glu Gln Val Phe Gly Thr Thr Gly Lys Phe Val Ile
                140                 145                 150
Phe Gly Ala Thr Ser Leu Gln Asn Thr Gly Ala Met Leu Ser Tyr
                155                 160                 165
Leu Phe Ile Val Lys Asn Glu Leu Pro Ser Ala Ile Lys Phe Leu
                170                 175                 180
Met Gly Lys Glu Glu Thr Phe Ser Ala Trp Tyr Val Asp Gly Arg
                185                 190                 195
Val Leu Val Val Ile Val Thr Phe Gly Ile Ile Leu Pro Leu Cys
                200                 205                 210
Leu Leu Lys Asn Leu Gly Tyr Leu Gly Tyr Thr Ser Gly Phe Ser
                215                 220                 225
Leu Ser Cys Met Val Phe Phe Leu Ile Val Val Ile Tyr Lys Lys
                230                 235                 240
Phe Gln Ile Pro Cys Ile Val Pro Glu Leu Asn Ser Thr Ile Ser
                245                 250                 255
Ala Asn Ser Thr Asn Ala Asp Thr Cys Thr Pro Lys Tyr Val Thr
                260                 265                 270
Phe Asn Ser Lys Thr Val Tyr Ala Leu Pro Thr Ile Ala Phe Ala
                275                 280                 285
Phe Val Cys His Pro Ser Val Leu Pro Ile Tyr Ser Glu Leu Lys
                290                 295                 300
Asp Arg Ser Gln Lys Lys Met Gln Met Val Ser Asn Ile Ser Phe
                305                 310                 315
Phe Ala Met Phe Val Met Tyr Phe Leu Thr Ala Ile Phe Gly Tyr
```

```
                320                 325                 330
Leu Thr Phe Tyr Asp Asn Val Gln Ser Asp Leu Leu His Lys Tyr
                335                 340                 345
Gln Ser Lys Asp Asp Ile Leu Ile Leu Thr Val Arg Leu Ala Val
                350                 355                 360
Ile Val Ala Val Ile Leu Thr Val Pro Val Leu Phe Phe Thr Val
                365                 370                 375
Arg Ser Ser Leu Phe Glu Leu Ala Lys Lys Thr Lys Phe Asn Leu
                380                 385                 390
Cys Arg His Thr Val Val Thr Cys Ile Leu Leu Val Val Ile Asn
                395                 400                 405
Leu Leu Val Ile Phe Ile Pro Ser Met Lys Asp Ile Phe Gly Val
                410                 415                 420
Val Gly Val Thr Ser Ala Asn Met Leu Ile Phe Ile Leu Pro Ser
                425                 430                 435
Ser Leu Tyr Leu Lys Ile Thr Asp Gln Asp Gly Asp Lys Gly Thr
                440                 445                 450
Gln Arg Ile Trp Ala Ala Leu Phe Leu Gly Leu Gly Val Leu Phe
                455                 460                 465
Ser Leu Val Ser Ile Pro Leu Val Ile Tyr Asp Trp Ala Cys Ser
                470                 475                 480
Ser Ser Ser Asp Glu Gly His
                485
```

<210> 13
<211> 509
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4027693CD1

<400> 13

```
Met Glu Leu Lys Lys Ser Pro Asp Gly Gly Trp Gly Trp Val Ile
  1                 5                  10                  15
Val Phe Val Ser Phe Leu Thr Gln Phe Leu Cys Tyr Gly Ser Pro
                 20                  25                  30
Leu Ala Val Gly Val Leu Tyr Ile Glu Trp Leu Asp Ala Phe Gly
                 35                  40                  45
Glu Gly Lys Gly Lys Thr Ala Trp Val Gly Ser Leu Ala Ser Gly
                 50                  55                  60
Val Gly Leu Leu Ala Ser Pro Val Cys Ser Leu Cys Val Ser Ser
                 65                  70                  75
Phe Gly Ala Arg Pro Val Thr Ile Phe Ser Gly Phe Met Val Ala
                 80                  85                  90
Gly Gly Leu Met Leu Ser Ser Phe Ala Pro Asn Ile Tyr Phe Leu
                 95                 100                 105
Phe Phe Ser Tyr Gly Ile Val Val Gly Leu Gly Cys Gly Leu Leu
                110                 115                 120
Tyr Thr Ala Thr Val Thr Ile Thr Cys Gln Tyr Phe Asp Asp Arg
                125                 130                 135
Arg Gly Leu Ala Leu Gly Leu Ile Ser Thr Gly Ser Ser Val Gly
                140                 145                 150
Leu Phe Ile Tyr Ala Ala Leu Gln Arg Met Leu Val Glu Phe Tyr
                155                 160                 165
Gly Leu Asp Gly Cys Leu Leu Ile Val Gly Ala Leu Ala Leu Asn
                170                 175                 180
Ile Leu Ala Cys Gly Ser Leu Met Arg Pro Leu Gln Ser Ser Asp
                185                 190                 195
Cys Pro Leu Pro Lys Lys Ile Ala Pro Glu Asp Leu Pro Asp Lys
                200                 205                 210
Tyr Ser Ile Tyr Asn Glu Lys Gly Lys Asn Leu Glu Glu Asn Ile
```

```
                215                 220                 225
Asn Ile Leu Asp Lys Ser Tyr Ser Ser Glu Glu Lys Cys Arg Ile
                230                 235                 240
Thr Leu Ala Asn Gly Asp Trp Lys Gln Asp Ser Leu Leu His Lys
                245                 250                 255
Asn Pro Thr Val Thr His Thr Lys Glu Pro Glu Thr Tyr Lys Lys
                260                 265                 270
Lys Val Ala Glu Gln Thr Tyr Phe Cys Lys Gln Leu Ala Lys Arg
                275                 280                 285
Lys Trp Gln Leu Tyr Lys Asn Tyr Cys Gly Glu Thr Val Ala Leu
                290                 295                 300
Phe Lys Asn Lys Val Phe Ser Ala Leu Phe Ile Ala Ile Leu Leu
                305                 310                 315
Phe Asp Ile Gly Gly Phe Pro Pro Ser Leu Leu Met Glu Asp Val
                320                 325                 330
Ala Arg Ser Ser Asn Val Lys Glu Glu Glu Phe Ile Met Pro Leu
                335                 340                 345
Ile Ser Ile Ile Gly Ile Met Thr Ala Val Gly Lys Leu Leu Leu
                350                 355                 360
Gly Ile Leu Ala Asp Phe Lys Trp Ile Asn Thr Leu Tyr Leu Tyr
                365                 370                 375
Val Ala Thr Leu Ile Ile Met Gly Leu Ala Leu Cys Ala Ile Pro
                380                 385                 390
Phe Ala Lys Ser Tyr Val Thr Leu Ala Leu Leu Ser Gly Ile Leu
                395                 400                 405
Gly Phe Leu Thr Gly Asn Trp Ser Ile Phe Pro Tyr Val Thr Thr
                410                 415                 420
Lys Thr Val Gly Ile Glu Lys Leu Ala His Ala Tyr Gly Ile Leu
                425                 430                 435
Met Phe Phe Ala Gly Leu Gly Asn Ser Leu Gly Pro Pro Ile Val
                440                 445                 450
Gly Trp Phe Tyr Asp Trp Thr Gln Thr Tyr Asp Ile Ala Phe Tyr
                455                 460                 465
Phe Ser Gly Phe Cys Val Leu Leu Gly Gly Phe Ile Leu Leu Leu
                470                 475                 480
Ala Ala Leu Pro Ser Trp Asp Thr Cys Asn Lys Gln Leu Pro Lys
                485                 490                 495
Pro Ala Pro Thr Thr Phe Leu Tyr Lys Val Ala Ser Asn Val
                500                 505
```

<210> 14
<211> 1232
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472030CD1

<400> 14

```
Met Val Tyr Ser Gly Asn Ala Glu Met Phe Asn Ile Gln Lys Ser
  1               5                  10                  15
Thr Ala Leu Ile Thr Ala Glu Glu Gln Pro Lys Leu Arg Lys Glu
                 20                  25                  30
Ala Val Gly Ser Ile Glu Ile Phe Arg Phe Ala Asp Gly Leu Asp
                 35                  40                  45
Ile Thr Leu Met Ile Leu Gly Ile Leu Thr Ser Leu Phe Asn Gly
                 50                  55                  60
Ala Cys Leu Pro Leu Met Pro Leu Cys Ile Gly Glu Met Ser Asp
                 65                  70                  75
Asn Leu Ile Ser Gly Cys Leu Val His Thr Asn Thr Thr Asn Tyr
                 80                  85                  90
Gln Asn Cys Thr Gln Ser Gln Glu Lys Leu Asn Glu Asp Met Thr
```

```
                95                  100                 105
Leu Leu Thr Leu Tyr Tyr Val Gly Ile Gly Val Ala Ala Leu Ile
            110                 115                 120
Phe Gly Tyr Ile Gln Ile Ser Leu Trp Ile Ile Thr Ala Ala Arg
        125                 130                 135
Gln Thr Lys Arg Ile Arg Lys Gln Phe Phe His Ser Val Leu Ala
    140                 145                 150
Gln Asp Ile Gly Trp Phe Asp Ser Cys Asp Ile Gly Glu Leu Asn
155                 160                 165
Thr Arg Met Thr Asp Asp Ile Asp Lys Ile Ser Asp Gly Ile Gly
            170                 175                 180
Asp Lys Ile Ala Leu Leu Phe Gln Asn Met Ser Thr Phe Ser Ile
        185                 190                 195
Gly Leu Ala Val Gly Leu Val Lys Gly Trp Lys Leu Thr Leu Val
    200                 205                 210
Thr Leu Ser Thr Ser Pro Leu Ile Met Ala Ser Ala Ala Ala Cys
215                 220                 225
Ser Arg Met Val Ile Ser Leu Thr Ser Lys Glu Leu Ser Ala Tyr
            230                 235                 240
Ser Lys Ala Gly Ala Val Ala Glu Glu Val Leu Ser Ser Ile Arg
        245                 250                 255
Thr Val Ile Ala Phe Arg Ala Gln Glu Lys Glu Leu Gln Arg Tyr
    260                 265                 270
Thr Gln Asn Leu Lys Asp Ala Lys Asp Phe Gly Ile Lys Arg Thr
275                 280                 285
Ile Ala Ser Lys Val Ser Leu Gly Ala Val Tyr Phe Phe Met Asn
            290                 295                 300
Gly Thr Tyr Gly Leu Ala Phe Trp Tyr Gly Thr Ser Leu Ile Leu
305                 310                 315
Asn Gly Glu Pro Gly Tyr Thr Ile Gly Thr Val Leu Ala Val Phe
            320                 325                 330
Phe Ser Val Ile His Ser Ser Tyr Cys Ile Gly Ala Ala Val Pro
        335                 340                 345
His Phe Glu Thr Phe Ala Ile Ala Arg Gly Ala Ala Phe His Ile
    350                 355                 360
Phe Gln Val Ile Asp Lys Lys Pro Ser Ile Gly Asn Phe Ser Thr
365                 370                 375
Ala Gly Tyr Lys Pro Glu Ser Ile Glu Gly Thr Val Glu Phe Lys
            380                 385                 390
Asn Val Ser Phe Asn Tyr Pro Ser Arg Pro Ser Ile Lys Ile Leu
        395                 400                 405
Lys Gly Leu Asn Leu Gly Ile Lys Ser Gly Glu Thr Val Ala Leu
    410                 415                 420
Val Gly Leu Asn Gly Ser Gly Lys Ser Thr Val Val Gln Leu Leu
425                 430                 435
Gln Arg Leu Tyr Asp Pro Asp Asp Gly Phe Ile Met Val Asp Glu
            440                 445                 450
Asn Asp Ile Arg Ala Leu Asn Val Arg His Tyr Arg Asp His Ile
        455                 460                 465
Gly Val Val Ser Gln Glu Pro Val Leu Phe Gly Thr Thr Ile Ser
    470                 475                 480
Asn Asn Ile Lys Tyr Gly Arg Asp Asp Val Thr Asp Glu Glu Met
485                 490                 495
Glu Arg Ala Ala Arg Glu Ala Asn Ala Tyr Asp Phe Ile Met Glu
            500                 505                 510
Phe Pro Asn Lys Phe Asn Thr Leu Val Gly Glu Lys Gly Ala Gln
        515                 520                 525
Met Ser Gly Gly Gln Lys Gln Arg Ile Ala Ile Ala Arg Ala Leu
    530                 535                 540
Val Arg Asn Pro Lys Ile Leu Ile Leu Asp Glu Ala Thr Ser Ala
545                 550                 555
Leu Asp Ser Glu Ser Lys Ser Ala Val Gln Ala Ala Leu Glu Lys
            560                 565                 570
```

```
Ala Ser Lys Gly Arg Thr Thr Ile Val Val Ala His Arg Leu Ser
                575                 580                 585
Thr Ile Arg Ser Ala Asp Leu Ile Val Thr Leu Lys Asp Gly Met
                590                 595                 600
Leu Ala Glu Lys Gly Ala His Ala Glu Leu Met Ala Lys Arg Gly
                605                 610                 615
Leu Tyr Tyr Ser Leu Val Met Ser Gln Asp Ile Lys Lys Ala Asp
                620                 625                 630
Glu Gln Met Glu Ser Met Thr Tyr Ser Thr Glu Arg Lys Thr Asn
                635                 640                 645
Ser Leu Pro Leu His Ser Val Lys Ser Ile Lys Ser Asp Phe Ile
                650                 655                 660
Asp Lys Ala Glu Glu Ser Thr Gln Ser Lys Glu Ile Ser Leu Pro
                665                 670                 675
Glu Val Ser Leu Leu Lys Ile Leu Lys Leu Asn Lys Pro Glu Trp
                680                 685                 690
Pro Phe Val Val Leu Gly Thr Leu Ala Ser Val Leu Asn Gly Thr
                695                 700                 705
Val His Pro Val Phe Ser Ile Ile Phe Ala Lys Ile Ile Thr Met
                710                 715                 720
Phe Gly Asn Asn Asp Lys Thr Thr Leu Lys His Asp Ala Glu Ile
                725                 730                 735
Tyr Ser Met Ile Phe Val Ile Leu Gly Val Ile Cys Phe Val Ser
                740                 745                 750
Tyr Phe Met Gln Asp Ile Ala Trp Phe Asp Glu Lys Glu Asn Ser
                755                 760                 765
Thr Gly Gly Leu Thr Thr Ile Leu Ala Ile Asp Ile Ala Gln Ile
                770                 775                 780
Gln Gly Ala Thr Gly Ser Arg Ile Gly Val Leu Thr Gln Asn Ala
                785                 790                 795
Thr Asn Met Gly Leu Ser Val Ile Ile Ser Phe Ile Tyr Gly Trp
                800                 805                 810
Glu Met Thr Phe Leu Ile Leu Ser Ile Ala Pro Val Leu Ala Val
                815                 820                 825
Thr Gly Met Ile Glu Thr Ala Ala Met Thr Gly Phe Ala Asn Lys
                830                 835                 840
Asp Lys Gln Glu Leu Lys His Ala Gly Lys Ile Ala Thr Glu Ala
                845                 850                 855
Leu Glu Asn Ile Arg Thr Ile Val Ser Leu Thr Arg Glu Lys Ala
                860                 865                 870
Phe Glu Gln Met Tyr Glu Glu Met Leu Gln Thr Gln His Arg Asn
                875                 880                 885
Thr Ser Lys Lys Ala Gln Ile Ile Gly Ser Cys Tyr Ala Phe Ser
                890                 895                 900
His Ala Phe Ile Tyr Phe Ala Tyr Ala Ala Gly Phe Arg Phe Gly
                905                 910                 915
Ala Tyr Leu Ile Gln Ala Gly Arg Met Thr Pro Glu Gly Met Phe
                920                 925                 930
Ile Val Phe Thr Ala Ile Ala Tyr Gly Ala Met Ala Ile Gly Glu
                935                 940                 945
Thr Leu Val Leu Ala Pro Glu Tyr Ser Lys Ala Lys Ser Gly Ala
                950                 955                 960
Ala His Leu Phe Ala Leu Leu Glu Lys Lys Pro Asn Ile Asp Ser
                965                 970                 975
Arg Ser Gln Glu Gly Lys Lys Pro Asp Thr Cys Glu Gly Asn Leu
                980                 985                 990
Glu Phe Arg Glu Val Ser Phe Phe Tyr Pro Cys Arg Pro Asp Val
                995                1000                1005
Phe Ile Leu Arg Gly Leu Ser Leu Ser Ile Glu Arg Gly Lys Thr
               1010                1015                1020
Val Ala Phe Val Gly Ser Ser Gly Cys Gly Lys Ser Thr Ser Val
               1025                1030                1035
Gln Leu Leu Gln Arg Leu Tyr Asp Pro Val Gln Gly Gln Val Leu
```

```
                1040                1045                1050
Phe Asp Gly Val Asp Ala Lys Glu Leu Asn Val Gln Trp Leu Arg
                1055                1060                1065
Ser Gln Ile Ala Ile Val Pro Gln Glu Pro Val Leu Phe Asn Cys
                1070                1075                1080
Ser Ile Ala Glu Asn Ile Ala Tyr Gly Asp Asn Ser Arg Val Val
                1085                1090                1095
Pro Leu Asp Glu Ile Lys Glu Ala Ala Asn Ala Ala Asn Ile His
                1100                1105                1110
Ser Phe Ile Glu Gly Leu Pro Glu Lys Tyr Asn Thr Gln Val Gly
                1115                1120                1125
Leu Lys Gly Ala Gln Leu Ser Gly Gly Gln Lys Gln Arg Leu Ala
                1130                1135                1140
Ile Ala Arg Ala Leu Leu Gln Lys Pro Lys Ile Leu Leu Leu Asp
                1145                1150                1155
Glu Ala Thr Ser Ala Leu Asp Asn Asp Ser Glu Lys Val Val Gln
                1160                1165                1170
His Ala Leu Asp Lys Ala Arg Thr Gly Arg Thr Cys Leu Val Val
                1175                1180                1185
Thr His Arg Leu Ser Ala Ile Gln Asn Ala Asp Leu Ile Val Val
                1190                1195                1200
Leu His Asn Gly Lys Ile Lys Glu Gln Gly Thr His Gln Glu Leu
                1205                1210                1215
Leu Arg Asn Arg Asp Ile Tyr Phe Lys Leu Val Asn Ala Gln Ser
                1220                1225                1230
Val Gln
```

<210> 15
<211> 759
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476089CD1

<400> 15

```
Met Thr Leu Pro Ala Ser Ser Val Pro His Ile Thr Val Glu Glu
  1               5                  10                  15
Glu Asp Gly Glu Ile Arg Leu Trp Ser Ser Val His Thr Gly Leu
                 20                  25                  30
Leu Gly Arg Val Thr Ala Glu Phe Arg Thr Val Ser Leu Thr Ala
                 35                  40                  45
Phe Ser Pro Glu Asp Tyr Gln Asn Val Ala Gly Thr Leu Glu Phe
                 50                  55                  60
Gln Pro Gly Glu Arg Tyr Lys Tyr Ile Phe Ile Asn Ile Thr Asp
                 65                  70                  75
Asn Ser Ile Pro Glu Leu Glu Lys Ser Phe Lys Val Glu Leu Leu
                 80                  85                  90
Asn Leu Glu Gly Gly Ala Ser Leu Gly Val Ala Ser Gln Ile Leu
                 95                 100                 105
Val Thr Ile Ala Ala Ser Asp His Ala His Gly Val Phe Glu Phe
                110                 115                 120
Ser Pro Glu Ser Leu Phe Val Ser Gly Thr Glu Pro Glu Asp Gly
                125                 130                 135
Tyr Ser Thr Val Thr Leu Asn Val Ile Arg His His Gly Thr Leu
                140                 145                 150
Ser Pro Val Thr Leu His Trp Asn Ile Asp Ser Asp Pro Asp Gly
                155                 160                 165
Asp Leu Ala Phe Thr Ser Gly Asn Ile Thr Phe Glu Ile Gly Gln
                170                 175                 180
Thr Ser Ala Asn Ile Thr Val Glu Ile Leu Pro Asp Glu Asp Pro
```

```
                185                 190                 195
Glu Leu Asp Lys Ala Phe Ser Val Ser Val Leu Ser Val Ser Ser
                200                 205                 210
Gly Ser Leu Gly Ala His Ile Asn Ala Thr Leu Thr Val Leu Ala
                215                 220                 225
Ser Asp Asp Pro Tyr Gly Ile Phe Ile Phe Ser Glu Lys Asn Arg
                230                 235                 240
Pro Val Lys Val Glu Glu Ala Thr Gln Asn Ile Thr Leu Ser Ile
                245                 250                 255
Ile Arg Leu Lys Gly Leu Met Gly Lys Val Leu Val Ser Tyr Ala
                260                 265                 270
Thr Leu Asp Asp Met Glu Lys Pro Pro Tyr Phe Pro Pro Asn Leu
                275                 280                 285
Ala Arg Ala Thr Gln Gly Arg Asp Tyr Ile Pro Ala Ser Gly Phe
                290                 295                 300
Ala Leu Phe Gly Ala Asn Gln Ser Glu Ala Thr Ile Ala Ile Ser
                305                 310                 315
Ile Leu Asp Asp Asp Glu Pro Glu Arg Ser Glu Ser Val Phe Ile
                320                 325                 330
Glu Leu Leu Asn Ser Thr Leu Val Ala Lys Val Gln Ser Arg Ser
                335                 340                 345
Ile Pro Asn Ser Pro Arg Leu Gly Pro Lys Val Glu Thr Ile Ala
                350                 355                 360
Gln Leu Ile Ile Ile Ala Asn Asp Asp Ala Phe Gly Thr Leu Gln
                365                 370                 375
Leu Ser Ala Pro Ile Val Arg Val Ala Glu Asn His Val Gly Pro
                380                 385                 390
Ile Ile Asn Val Thr Arg Thr Gly Gly Ala Phe Ala Asp Val Ser
                395                 400                 405
Val Lys Phe Lys Ala Val Pro Ile Thr Ala Ile Ala Gly Glu Asp
                410                 415                 420
Tyr Ser Ile Ala Ser Ser Asp Val Val Leu Leu Glu Gly Glu Thr
                425                 430                 435
Ser Lys Ala Val Pro Ile Tyr Val Ile Asn Asp Ile Tyr Pro Glu
                440                 445                 450
Leu Glu Glu Ser Phe Leu Val Gln Leu Met Asn Glu Thr Thr Gly
                455                 460                 465
Gly Ala Arg Leu Gly Ala Leu Thr Glu Ala Val Ile Ile Ile Glu
                470                 475                 480
Ala Ser Asp Asp Pro Tyr Gly Leu Phe Gly Phe Gln Ile Thr Lys
                485                 490                 495
Leu Ile Val Glu Glu Pro Glu Phe Asn Ser Val Lys Val Asn Leu
                500                 505                 510
Pro Ile Ile Arg Asn Ser Gly Thr Leu Gly Asn Val Thr Val Gln
                515                 520                 525
Trp Val Ala Thr Ile Asn Gly Gln Leu Ala Thr Gly Asp Leu Arg
                530                 535                 540
Val Val Ser Gly Asn Val Thr Phe Ala Pro Gly Glu Thr Ile Gln
                545                 550                 555
Thr Leu Leu Leu Glu Val Leu Ala Asp Asp Val Pro Glu Ile Glu
                560                 565                 570
Glu Val Ile Gln Val Gln Leu Thr Asp Ala Ser Gly Gly Gly Thr
                575                 580                 585
Ile Gly Leu Asp Arg Ile Ala Asn Ile Ile Ile Pro Ala Asn Asp
                590                 595                 600
Asp Pro Tyr Gly Thr Val Ala Phe Ala Gln Met Val Tyr Arg Val
                605                 610                 615
Gln Glu Pro Leu Glu Arg Ser Ser Cys Ala Asn Ile Thr Val Arg
                620                 625                 630
Arg Ser Gly Gly His Phe Gly Arg Leu Leu Leu Phe Tyr Ser Thr
                635                 640                 645
Ser Asp Ile Asp Val Val Ala Leu Ala Met Glu Glu Gly Gln Asp
                650                 655                 660
```

```
Leu Leu Ser Tyr Tyr Glu Ser Pro Ile Gln Gly Val Pro Asp Pro
                665                 670                 675
Leu Trp Arg Thr Trp Met Asn Val Ser Ala Val Gly Glu Pro Leu
                680                 685                 690
Tyr Thr Cys Ala Thr Leu Cys Leu Lys Glu Gln Ala Cys Ser Ala
                695                 700                 705
Phe Ser Phe Phe Ser Ala Ser Glu Gly Pro Gln Cys Phe Trp Met
                710                 715                 720
Thr Ser Trp Ile Ser Pro Ala Val Asn Asn Ser Asp Phe Trp Thr
                725                 730                 735
Tyr Arg Lys Asn Met Thr Arg Val Ala Ser Leu Leu Val Val Arg
                740                 745                 750
Leu Trp Leu Gly Val Thr Met Ser Leu
                755
```

<210> 16
<211> 283
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6428177CD1

<400> 16

```
Met Pro His Arg Lys Glu Arg Pro Ser Gly Ser Ser Leu His Thr
  1                 5                  10                  15
His Gly Ser Thr Gly Thr Ala Glu Gly Gly Asn Met Ser Arg Leu
                 20                  25                  30
Ser Leu Thr Arg Ser Pro Val Ser Pro Leu Ala Ala Gln Gly Ile
                 35                  40                  45
Pro Leu Pro Ala Gln Leu Thr Lys Ser Asn Ala Pro Val His Ile
                 50                  55                  60
Asp Val Gly Gly His Met Tyr Thr Ser Ser Leu Ala Thr Leu Thr
                 65                  70                  75
Lys Tyr Pro Asp Ser Arg Ile Ser Arg Leu Phe Asn Gly Thr Glu
                 80                  85                  90
Pro Ile Val Leu Asp Ser Leu Lys Gln His Tyr Phe Ile Asp Arg
                 95                 100                 105
Asp Gly Glu Ile Phe Arg Tyr Val Leu Ser Phe Leu Arg Thr Ser
                110                 115                 120
Lys Leu Leu Leu Pro Asp Asp Phe Lys Asp Phe Ser Leu Leu Tyr
                125                 130                 135
Glu Glu Ala Arg Tyr Tyr Gln Leu Gln Pro Met Val Arg Glu Leu
                140                 145                 150
Glu Arg Trp Gln Gln Glu Gln Glu Gln Arg Arg Arg Ser Arg Ala
                155                 160                 165
Cys Asp Cys Leu Val Val Arg Val Thr Pro Asp Leu Gly Glu Arg
                170                 175                 180
Ile Ala Leu Ser Gly Glu Lys Ala Leu Ile Glu Glu Val Phe Pro
                185                 190                 195
Glu Thr Gly Asp Val Met Cys Asn Ser Val Asn Ala Gly Trp Asn
                200                 205                 210
Gln Asp Pro Thr His Val Ile Arg Phe Pro Leu Asn Gly Tyr Cys
                215                 220                 225
Arg Leu Asn Ser Val Gln Val Leu Glu Arg Leu Phe Gln Arg Gly
                230                 235                 240
Phe Ser Val Ala Ala Ser Cys Gly Gly Gly Val Asp Ser Ser Gln
                245                 250                 255
Phe Ser Glu Tyr Val Leu Cys Arg Glu Glu Arg Arg Pro Gln Pro
                260                 265                 270
Thr Pro Thr Ala Val Arg Ile Lys Gln Glu Pro Leu Asp
                275                 280
```

<210> 17
<211> 1129
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477243CD1

<400> 17
Met Phe Arg Arg Ser Leu Asn Arg Phe Cys Ala Gly Glu Glu Lys
1 5 10 15
Arg Val Gly Thr Arg Thr Val Phe Val Gly Asn His Pro Val Ser
20 25 30
Glu Thr Glu Ala Tyr Ile Ala Gln Arg Phe Cys Asp Asn Arg Ile
35 40 45
Val Ser Ser Lys Tyr Thr Leu Trp Asn Phe Leu Pro Lys Asn Leu
50 55 60
Phe Glu Gln Phe Arg Arg Ile Ala Asn Phe Tyr Phe Leu Ile Ile
65 70 75
Phe Leu Val Gln Val Thr Val Asp Thr Pro Thr Ser Pro Val Thr
80 85 90
Ser Gly Leu Pro Leu Phe Phe Val Ile Thr Val Thr Ala Ile Lys
95 100 105
Gln Gly Tyr Glu Asp Cys Leu Arg His Arg Ala Asp Asn Glu Val
110 115 120
Asn Lys Ser Thr Val Tyr Ile Ile Glu Asn Ala Lys Arg Val Arg
125 130 135
Lys Glu Ser Glu Lys Ile Lys Val Gly Asp Val Val Glu Val Gln
140 145 150
Ala Asp Glu Thr Phe Pro Cys Asp Leu Ile Leu Leu Ser Ser Cys
155 160 165
Thr Thr Asp Gly Thr Cys Tyr Val Thr Thr Ala Ser Leu Asp Gly
170 175 180
Glu Ser Asn Cys Lys Thr His Tyr Ala Val Arg Asp Thr Ile Ala
185 190 195
Leu Cys Thr Ala Glu Ser Ile Asp Thr Leu Arg Ala Ala Ile Glu
200 205 210
Cys Glu Gln Pro Gln Pro Asp Leu Tyr Lys Phe Val Gly Arg Ile
215 220 225
Asn Ile Tyr Ser Asn Ser Leu Glu Ala Val Ala Arg Ser Leu Gly
230 235 240
Pro Glu Asn Leu Leu Leu Lys Gly Ala Thr Leu Lys Asn Thr Glu
245 250 255
Lys Ile Tyr Gly Val Ala Val Tyr Thr Gly Met Glu Thr Lys Met
260 265 270
Ala Leu Asn Tyr Gln Gly Lys Ser Gln Lys Arg Ser Ala Val Glu
275 280 285
Lys Ser Ile Asn Ala Phe Leu Ile Val Tyr Leu Phe Ile Leu Leu
290 295 300
Thr Lys Ala Ala Val Cys Thr Thr Leu Lys Tyr Val Trp Gln Ser
305 310 315
Thr Pro Tyr Asn Asp Glu Pro Trp Tyr Asn Gln Lys Thr Gln Lys
320 325 330
Glu Arg Glu Thr Leu Lys Val Leu Lys Met Phe Thr Asp Phe Leu
335 340 345
Ser Phe Met Val Leu Phe Asn Phe Ile Ile Pro Val Ser Met Tyr
350 355 360
Val Thr Val Glu Met Gln Lys Phe Leu Gly Ser Phe Phe Ile Ser
365 370 375
Trp Asp Lys Asp Phe Tyr Asp Glu Glu Ile Asn Glu Gly Ala Leu
380 385 390
Val Asn Thr Ser Asp Leu Asn Glu Glu Leu Gly Gln Val Asp Tyr

```
                395                 400                 405
Val Phe Thr Asp Lys Thr Gly Thr Leu Thr Glu Asn Ser Met Glu
                410                 415                 420
Phe Ile Glu Cys Cys Ile Asp Gly His Lys Tyr Lys Gly Val Thr
                425                 430                 435
Gln Glu Val Asp Gly Leu Ser Gln Thr Asp Gly Thr L u Thr Tyr
                440                 445                 450
Phe Asp Lys Val Asp Lys Asn Arg Glu Glu Leu Phe Leu Arg Ala
                455                 460                 465
Leu Cys Leu Cys His Thr Val Glu Ile Lys Thr Asn Asp Ala Val
                470                 475                 480
Asp Gly Ala Thr Glu Ser Ala Glu Leu Thr Tyr Ile Ser Ser Ser
                485                 490                 495
Pro Asp Glu Ile Ala Leu Val Lys Gly Ala Lys Arg Tyr Gly Phe
                500                 505                 510
Thr Phe Leu Gly Asn Arg Asn Gly Tyr Met Arg Val Glu Asn Gln
                515                 520                 525
Arg Lys Glu Ile Glu Glu Tyr Glu Leu Leu His Thr Leu Asn Phe
                530                 535                 540
Asp Ala Val Arg Arg Arg Met Ser Val Ile Val Lys Thr Gln Glu
                545                 550                 555
Gly Asp Ile Leu Leu Phe Cys Lys Gly Ala Asp Ser Ala Val Phe
                560                 565                 570
Pro Arg Val Gln Asn His Glu Ile Glu Leu Thr Lys Val His Val
                575                 580                 585
Glu Arg Asn Ala Met Asp Gly Tyr Arg Thr Leu Cys Val Ala Phe
                590                 595                 600
Lys Glu Ile Ala Pro Asp Asp Tyr Glu Arg Ile Asn Arg Gln Leu
                605                 610                 615
Ile Glu Ala Lys Met Ala Leu Gln Asp Arg Glu Glu Lys Met Glu
                620                 625                 630
Lys Val Phe Asp Asp Ile Glu Thr Asn Met Asn Leu Ile Gly Ala
                635                 640                 645
Thr Ala Val Glu Asp Lys Leu Gln Asp Gln Ala Ala Glu Thr Ile
                650                 655                 660
Glu Ala Leu His Ala Ala Gly Leu Lys Val Trp Val Leu Thr Gly
                665                 670                 675
Asp Lys Met Glu Thr Ala Lys Ser Thr Cys Tyr Ala Cys Arg Leu
                680                 685                 690
Phe Gln Thr Asn Thr Glu Leu Leu Glu Leu Thr Thr Lys Thr Ile
                695                 700                 705
Glu Glu Ser Glu Arg Lys Glu Asp Arg Leu His Glu Leu Leu Ile
                710                 715                 720
Glu Tyr Arg Lys Lys Leu Leu His Glu Phe Pro Lys Ser Thr Arg
                725                 730                 735
Ser Phe Lys Lys Ala Trp Thr Glu His Gln Glu Tyr Gly Leu Ile
                740                 745                 750
Ile Asp Gly Ser Thr Leu Ser Leu Ile Leu Asn Ser Ser Gln Asp
                755                 760                 765
Ser Ser Ser Asn Asn Tyr Lys Ser Ile Phe Leu Gln Ile Cys Met
                770                 775                 780
Lys Cys Thr Ala Val Leu Cys Cys Arg Met Ala Pro Leu Gln Lys
                785                 790                 795
Ala Gln Ile Val Arg Met Val Lys Asn Leu Lys Gly Ser Pro Ile
                800                 805                 810
Thr Leu Ser Ile Gly Asp Gly Ala Asn Asp Val Ser Met Ile Leu
                815                 820                 825
Glu Ser His Val Gly Ile Gly Ile Lys Gly Lys Glu Gly Arg Gln
                830                 835                 840
Ala Ala Arg Asn Ser Asp Tyr Ser Val Pro Lys Phe Lys His Leu
                845                 850                 855
Lys Lys Leu Leu Leu Ala His Gly His Leu Tyr Tyr Val Arg Ile
                860                 865                 870
```

```
Ala His Leu Val Gln Tyr Phe Phe Tyr Lys Asn Leu Cys Phe Ile
                875                 880                 885
Leu Pro Gln Phe Leu Tyr Gln Phe Phe Cys Gly Phe Ser Gln Gln
                890                 895                 900
Pro Leu Tyr Asp Ala Ala Tyr Leu Thr Met Tyr Asn Ile Cys Phe
                905                 910                 915
Thr Ser Leu Pro Ile Leu Ala Tyr Ser Leu Leu Glu Gln His Ile
                920                 925                 930
Asn Ile Asp Thr Leu Thr Ser Asp Pro Arg Leu Tyr Met Lys Ile
                935                 940                 945
Ser Gly Asn Ala Met Leu Gln Leu Gly Pro Phe Leu Tyr Trp Thr
                950                 955                 960
Phe Leu Ala Ala Phe Glu Gly Thr Val Phe Phe Phe Gly Thr Tyr
                965                 970                 975
Phe Leu Phe Gln Thr Ala Ser Leu Glu Glu Asn Gly Lys Val Tyr
                980                 985                 990
Gly Asn Trp Thr Phe Gly Thr Ile Val Phe Thr Val Leu Val Phe
                995                1000                1005
Thr Val Thr Leu Lys Leu Ala Leu Asp Thr Arg Phe Trp Thr Trp
               1010                1015                1020
Ile Asn His Phe Val Ile Trp Gly Ser Leu Ala Phe Tyr Val Phe
               1025                1030                1035
Phe Ser Phe Phe Trp Gly Gly Ile Ile Trp Pro Phe Leu Lys Gln
               1040                1045                1050
Gln Arg Met Tyr Phe Val Phe Ala Gln Met Leu Ser Ser Val Ser
               1055                1060                1065
Thr Trp Leu Ala Ile Ile Leu Leu Ile Phe Ile Ser Leu Phe Pro
               1070                1075                1080
Glu Ile Leu Leu Ile Val Leu Lys Asn Val Arg Arg Arg Ser Ala
               1085                1090                1095
Arg Arg Asn Leu Ser Cys Arg Arg Ala Ser Asp Ser Leu Ser Ala
               1100                1105                1110
Arg Pro Ser Val Arg Pro Leu Leu Leu Arg Thr Phe Ser Asp Glu
               1115                1120                1125
Ser Asn Val Leu
```

<210> 18
<211> 648
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473042CD1

<400> 18

```
Met Ser Arg Lys Ala Ser Glu Asn Val Glu Tyr Thr Leu Arg Ser
  1               5                  10                  15
Leu Ser Ser Leu Met Gly Glu Arg Arg Arg Lys Gln Pro Glu Pro
                20                  25                  30
Asp Ala Ala Ser Ala Ala Gly Glu Cys Ser Leu Leu Ala Ala Ala
                35                  40                  45
Glu Ser Ser Thr Ser Leu Gln Ser Ala Gly Ala Gly Gly Gly Gly
                50                  55                  60
Val Gly Asp Leu Glu Arg Ala Ala Arg Arg Gln Phe Gln Gln Asp
                65                  70                  75
Glu Thr Pro Ala Phe Val Tyr Val Val Ala Val Phe Ser Ala Leu
                80                  85                  90
Gly Gly Phe Leu Phe Gly Tyr Asp Thr Gly Val Val Ser Gly Ala
                95                 100                 105
Met Leu Leu Leu Lys Arg Gln Leu Ser Leu Asp Ala Leu Trp Gln
               110                 115                 120
```

```
Glu Leu Leu Val Ser Ser Thr Val Gly Ala Ala Ala Val Ser Ala
                125                 130                 135
Leu Ala Gly Gly Ala Leu Asn Gly Val Phe Gly Arg Arg Ala Ala
            140                 145                 150
Ile Leu Leu Ala Ser Ala Leu Phe Thr Ala Gly Ser Ala Val Leu
        155                 160                 165
Ala Ala Ala Asn Asn Lys Glu Thr Leu Leu Ala Gly Arg Leu Val
    170                 175                 180
Val Gly Leu Gly Ile Gly Ile Ala Ser Met Thr Val Pro Val Tyr
185                 190                 195
Ile Ala Glu Val Ser Pro Pro Asn Leu Arg Gly Arg Leu Val Thr
            200                 205                 210
Ile Asn Thr Leu Phe Ile Thr Gly Gly Gln Phe Phe Ala Ser Val
        215                 220                 225
Val Asp Gly Ala Phe Ser Tyr Leu Gln Lys Asp Gly Trp Arg Tyr
    230                 235                 240
Met Leu Gly Leu Ala Val Val Pro Ala Val Ile Gln Phe Phe Gly
245                 250                 255
Phe Leu Phe Leu Pro Glu Ser Pro Arg Trp Leu Ile Gln Lys Gly
            260                 265                 270
Gln Thr Gln Lys Ala Arg Arg Ile Leu Ser Gln Met Arg Gly Asn
        275                 280                 285
Gln Thr Ile Asp Glu Glu Tyr Asp Ser Ile Lys Asn Asn Ile Glu
    290                 295                 300
Glu Glu Glu Lys Glu Val Gly Ser Ala Gly Pro Val Ile Cys Arg
305                 310                 315
Met Leu Ser Tyr Pro Gln Thr Arg Arg Ala Leu Ile Val Gly Cys
            320                 325                 330
Gly Leu Gln Met Phe Gln Gln Leu Ser Gly Ile Asn Thr Ile Met
        335                 340                 345
Tyr Tyr Ser Ala Thr Ile Leu Gln Met Ser Gly Val Glu Asp Asp
    350                 355                 360
Arg Leu Ala Ile Trp Leu Ala Ser Val Thr Ala Phe Thr Asn Phe
365                 370                 375
Ile Phe Thr Leu Val Gly Val Trp Leu Val Glu Lys Val Gly Arg
            380                 385                 390
Arg Lys Leu Thr Phe Gly Ser Leu Ala Gly Thr Thr Val Ala Leu
        395                 400                 405
Ile Ile Leu Ala Leu Gly Phe Val Leu Ser Ala Gln Val Ser Pro
    410                 415                 420
Arg Ile Thr Phe Lys Pro Ile Ala Pro Ser Gly Gln Asn Ala Thr
425                 430                 435
Cys Thr Arg Tyr Ser Tyr Cys Asn Glu Cys Met Leu Asp Pro Asp
            440                 445                 450
Cys Gly Phe Cys Tyr Lys Met Asn Lys Ser Thr Val Ile Asp Ser
        455                 460                 465
Ser Cys Val Pro Val Asn Lys Ala Ser Thr Asn Glu Ala Ala Trp
    470                 475                 480
Gly Arg Cys Glu Asn Glu Thr Lys Phe Lys Thr Glu Asp Ile Phe
485                 490                 495
Trp Ala Tyr Asn Phe Cys Pro Thr Pro Tyr Ser Trp Thr Ala Leu
            500                 505                 510
Leu Gly Leu Ile Leu Tyr Leu Val Phe Phe Ala Pro Gly Met Gly
        515                 520                 525
Pro Met Pro Trp Thr Val Asn Ser Glu Ile Tyr Pro Leu Trp Ala
    530                 535                 540
Arg Ser Thr Gly Asn Ala Cys Ser Ser Gly Ile Asn Trp Ile Phe
545                 550                 555
Asn Val Leu Val Ser Leu Thr Phe Leu His Thr Ala Glu Tyr Leu
            560                 565                 570
Thr Tyr Tyr Gly Ala Phe Phe Leu Tyr Ala Gly Phe Ala Ala Val
        575                 580                 585
Gly Leu Leu Phe Ile Tyr Gly Cys Leu Pro Glu Thr Lys Gly Lys
```

```
                590                 595                 600
Lys Leu Glu Glu Ile Glu Ser Leu Phe Asp Asn Arg Leu Cys Thr
                605                 610                 615
Cys Gly Thr Ser Asp Ser Asp Glu Gly Arg Tyr Ile Glu Tyr Ile
                620                 625                 630
Arg Val Lys Gly Ser Asn Tyr His Leu Ser Asp Asn Asp Ala Ser
                635                 640                 645
Asp Val Glu
```

<210> 19
<211> 545
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482060CD1

<400> 19

```
Met Thr Phe Gly Arg Ser Gly Ala Ala Ser Val Val Leu Asn Val
  1               5                  10                  15
Gly Gly Ala Arg Tyr Ser Leu Ser Arg Glu Leu Leu Lys Asp Phe
                 20                  25                  30
Pro Leu Arg Arg Val Ser Arg Leu His Gly Cys Arg Ser Glu Arg
                 35                  40                  45
Asp Val Leu Glu Val Cys Asp Asp Tyr Asp Arg Glu Arg Asn Glu
                 50                  55                  60
Tyr Phe Phe Asp Arg His Ser Glu Ala Phe Gly Phe Ile Leu Leu
                 65                  70                  75
Tyr Val Arg Gly His Gly Lys Leu Arg Phe Ala Pro Arg Met Cys
                 80                  85                  90
Glu Leu Ser Phe Tyr Asn Glu Met Ile Tyr Trp Gly Leu Glu Gly
                 95                 100                 105
Ala His Leu Glu Tyr Cys Cys Gln Arg Arg Leu Asp Asp Arg Met
                110                 115                 120
Ser Asp Thr Tyr Thr Phe Tyr Ser Ala Asp Glu Pro Gly Val Leu
                125                 130                 135
Gly Arg Asp Glu Ala Arg Pro Gly Ala Arg Gly Gly Ser Leu Gln
                140                 145                 150
Ala Leu Ala Gly Ala His Ala Ala Asp Leu Arg Gly Ala His Ile
                155                 160                 165
Leu Ala Ser Val Ser Val Val Phe Val Ile Val Ser Met Val Val
                170                 175                 180
Leu Cys Ala Ser Thr Leu Pro Asp Trp Arg Asn Ala Ala Ala Asp
                185                 190                 195
Asn Arg Ser Leu Asp Asp Arg Ser Arg Ile Ile Glu Ala Ile Cys
                200                 205                 210
Ile Gly Trp Phe Thr Ala Glu Cys Ile Val Arg Phe Ile Val Ser
                215                 220                 225
Lys Asn Lys Cys Glu Phe Val Lys Arg Pro Leu Asn Ile Ile Asp
                230                 235                 240
Leu Leu Ala Ile Thr Pro Tyr Tyr Ile Ser Val Leu Met Thr Val
                245                 250                 255
Phe Thr Gly Glu Asn Ser Gln Leu Gln Arg Ala Gly Val Thr Leu
                260                 265                 270
Arg Val Leu Arg Met Met Arg Ile Phe Trp Val Ile Lys Leu Ala
                275                 280                 285
Arg His Phe Ile Gly Leu Gln Thr Leu Gly Leu Thr Leu Lys Arg
                290                 295                 300
Cys Tyr Arg Glu Met Val Met Leu Leu Val Phe Ile Cys Val Ala
                305                 310                 315
Met Ala Ile Phe Ser Ala Leu Ser Gln Leu Leu Glu His Gly Leu
```

```
                320                 325                 330
Asp Leu Glu Thr Ser Asn Lys Asp Phe Thr Ser Ile Pro Ala Ala
                335                 340                 345
Cys Trp Trp Val Ile Ile Ser Met Thr Thr Val Gly Tyr Gly Asp
                350                 355                 360
Met Tyr Pro Ile Thr Val Pro Gly Arg Ile Leu Gly Gly Val Cys
                365                 370                 375
Val Val Ser Gly Ile Val Leu Leu Ala Leu Pro Ile Thr Phe Ile
                380                 385                 390
Tyr His Ser Phe Val Gln Cys Tyr His Glu Leu Lys Phe Arg Ser
                395                 400                 405
Ala Arg Ser Ile Cys Leu Thr Ser Val Thr Ser Val Leu Gly Thr
                410                 415                 420
Val Gly Tyr Thr Glu Met Thr Ile Asn Gly Pro Cys Pro Asp Ala
                425                 430                 435
Leu Arg Asp Pro Cys Thr Cys Lys Lys Pro Leu Lys Thr His Ser
                440                 445                 450
Gly Val Leu Tyr Lys Ala Met Ala Asp Leu Trp Gln Ser Leu Glu
                455                 460                 465
Gly Gly Pro Pro Val Glu Gln Leu Pro Pro Asp Pro Leu Thr Arg
                470                 475                 480
Trp Cys Phe His Pro Ala Gly Ser Thr Leu Cys Gly Pro Ala Asn
                485                 490                 495
Ser Met Ala Val Ala Ser Pro Gly Ser Arg Pro Ala Ala Pro Gly
                500                 505                 510
Gly Gly Phe Leu Arg Thr Glu Ala Leu Val Leu Ile Val Ala Ala
                515                 520                 525
Gly Pro Val Asp Gly Leu Asn Cys Glu Asn His Pro Phe Arg Gly
                530                 535                 540
Gly Cys Lys Asp Phe
                545
```

<210> 20
<211> 262
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1578772CD1

<400> 20

```
Met Trp Gly Trp Glu Ala Leu Phe Leu Phe Cys Ser Cys Ser Ser
  1                 5                  10                  15
Phe Ser Leu Ala Gly Arg Pro Leu Leu Leu His Ser Gly Pro Val
                 20                  25                  30
Gly Ala Ala Val Ala Gly Ala Leu Leu Leu Leu Ser Ala Gln Gly
                 35                  40                  45
Cys Pro Gly Leu His Gln His Leu Gln His Ala Pro Gly Val Leu
                 50                  55                  60
Pro Asp Ala Gly Thr Ser Thr Thr Met Ala His Gln Pro Ser Gly
                 65                  70                  75
Leu Cys Cys Val Asp Gly His Leu Gly Gly Ser Ser Asp Pro Glu
                 80                  85                  90
Cys Gly Phe Gly Pro Gly Cys Gly Cys Gly Leu Leu His Asp Asp
                 95                 100                 105
Cys Gly Leu Pro His Pro Glu Leu Leu Gln Val Pro Gly Leu Cys
                110                 115                 120
Ile Leu Ser Tyr Pro Thr Pro Leu Tyr Phe Gly Thr Arg Gly Gln
                125                 130                 135
Phe Arg Cys Asn Leu Glu Trp His Leu Gly Leu Gly Glu Gly Glu
                140                 145                 150
Lys Glu Thr Ser Lys Pro Asp Gly Pro Met Val Ala Val Ala Glu
```

```
                155                 160                 165
Pro Val Arg Val Val Val Leu Asp Phe Ser Gly Val Thr Phe Ala
                170                 175                 180
Asp Ala Ala Gly Ala Arg Glu Val Val Gln Leu Ala Ser Arg Cys
                185                 190                 195
Arg Asp Ala Arg Ile Arg Leu Leu Leu Ala Gln Cys Asn Ala Leu
                200                 205                 210
Val Gln Gly Thr Leu Thr Arg Val Gly Leu Leu Asp Arg Val Thr
                215                 220                 225
Pro Asp Gln Leu Phe Val Ser Val Gln Asp Ala Ala Ala Tyr Ala
                230                 235                 240
Leu Gly Ser Leu Val Arg Gly Ser Ser Thr Arg Ser Gly Ser Gln
                245                 250                 255
Glu Ala Leu Gly Cys Gly Lys
                260
```

<210> 21
<211> 1373
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1626101CB1

<400> 21
cacgcgcggc ctggcggcgg cggccactct aaccagcgca aaatgtccct ggaacaggag 60
gaggaaacgc aacctgggcg gctcctagga cgcagagacg ccgtccccgc cttcattgag 120
cccaacgtgc gcttctggat caccgagcgc caatccttta ttcgacgatt tcttcaatgg 180
acagaattat tagatcctac aaatgtgttc atttcagttg aaagtataga aaactcgagg 240
caactattgt gcacaaatga agatgtttcc agccctgcct cggcggacca aaggatacag 300
gaagcttgga agcggagtct tgcaacagtg catcccgaca gcagcaacct gatccccaag 360
ctttttcgac ctgcagcgtt cctgcctttc atggcgccca cggtattttt gtcaatgacg 420
ccactgaaag ggatcaagtc cgtgatttta cctcaggttt tcctctgtgc ctacatggca 480
gcgttcaaca gcatcaatgg aaacagaagt tacacttgta agccactaga aagatcatta 540
ctaatggcgg gagccgttgc ttcttcaact ttcttaggag taatccctca gtttgtccag 600
atgaagtatg gcctgactgg cccttggatt aaaagactct tacctgtgat cttcctcgtg 660
caagccagtg gaatgaatgt ctacatgtcc cgaagtcttg aatccattaa ggggattgcg 720
gtcatggaca aggaaggcaa tgtcctgggt cattccagaa ttgctgggac aaaggctgtt 780
agagaaacgc tagcatccag aatagtgctg tttgggacct cagctctgat tcctgaagtc 840
ttcacctact tttttaaaag gacccagtat ttcaggaaaa acccagggtc attgtggatt 900
ttgaaactgt cttgtactgt cctggcaatg ggactgatgg tgccattttc ttttagtata 960
tttccacaga ttggacagat acagtactgt agtcttgaag agaaaattca gtctccaaca 1020
gaagaaacag aaatctttta tcacagaggg gtgtaggcgt gagttttagg tgaatttatg 1080
tggttcctgc ttgaaaacct tcccctctcc aggttcggtt tagagaactt tgccacaggt 1140
cttctgggga ccccagaggt gtctgtgctg acaaggcgac ttcagattcc atactgagat 1200
cgttcccagg ctggcgtctc tggggttttt aaggctggct ggagaagaca gtgggagggt 1260
gccccgtctg acacccctgg ggttgctgag ggaacggttg gagtggggat cggcctgcga 1320
aaggatactg tgaaatcact aattaactaa taaacctgtc tcaagttgag gaa 1373

<210> 22
<211> 3231
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2907828CB1

<400> 22
ttcggcggct gcggcggctg caacagcttc gggctcgggg ttttggcggc ggcgccggcg 60
ggctaggctg cgcggtgcgg accccggcgc gcggtccggg ttgctggggc ggcgcgtaag 120
atgcctctaa tggaggagtt tctgagcagc acccctggcc cagtggcttt gaaagggagc 180

tcaaaccaga gactatttca agccctggat atcatatcct gagggccaca ggagaagaga 240
acatggctgt gagtttggat gacgacgtgc cgctcatcct gaccttggat gagggtggca 300
gtgccccact ggctccctcc aacggcctgg gccaagaaga gctacctagc aaaaatggcg 360
gcagctatgc catccacgac tcccaggccc ccagtctcag ctctgggggt gagagttccc 420
cctccagccc cgcacacaac tgggagatga attaccaaga ggcagcaatc tacctccagg 480
aaggcgagaa caacgacaag ttcttcaccc accccaagga tgccaaggcg ctggcggcct 540
acctctttgc acacaatcac ctcttctacc tgatggagct ggccacggcc ctgctgctgc 600
tgctgctctc cctgtgcgag gcccccgccg tccccgcact ccggcttggc atctatgtcc 660
acgccaccct ggagctgttt gccctgatgg tggtagtgtt tgaactctgc atgaagttac 720
gctggctggg cctccacacc ttcatccggc acaagcggac catggtcaag acctcggtgc 780
tggtggtgca gtttgtcgag gccatcgtgg tgttggtacg gcagatgtcc catgtgcggg 840
tgacccgagc actgcgctgc attttcctgg tggactgtcg gtattgcggt ggcgtccggc 900
gcaacctgcg gcagatcttc cagtccctgc cgcccttcat ggacatcctc ctgctgctgc 960
tgttcttcat gatcatcttt gccatcctcg gtttctactt gttctccccc aacccttcag 1020
acccctactt cagcaccctg gagaacagca tcgtcagtct gtttgtcctt ctgaccacag 1080
ccaatttccc agatgtgatg atgccctcct actcccggaa cccctggtcc tgcgtcttct 1140
tcatcgtgta cctctccatc gagctgtatt tcatcatgaa cctgcttctg gctgtggtgt 1200
tcgacacctt caatgacatt gagaaacgca agttcaagtc tttgctactg cacaagcgaa 1260
ccgctatcca gcatgcctac cgcctgctca tcagccagag gaggcctgcc ggcatctcct 1320
acaggcagtt tgaaggcctc atgcgcttct acaagccccg gatgagtgcc agggagcgct 1380
atcttacctt caaggccctg aatcagaaca acacaccccc gctcagccta aaggactttt 1440
acgatatcta cgaagttgct gctttgaagt ggaaggccaa gaaaaacaga gagcactggt 1500
ttgatgagct tcccaggacg gcgctcctca tcttcaaagg tattaatatc cttgtgaagt 1560
ccaaggcctt ccagtatttc atgtacttgg tggtggcagt caacggggtc tggatcctcg 1620
tggagacatt tatgctgaaa ggtgggaact tcttctccaa gcacgtgccc tggagttacc 1680
tcgtctttct aactatctat ggggtggagc tgttcctgaa ggttgccggc ctgggccctg 1740
tggagtactt gtcttccgga tggaacttgt ttgacttctc cgtgacagtg ttcgccttcc 1800
tgggactgct ggcgctggcc ctcaacatgg agcccttcta tttcatcgtg gtcctgcgcc 1860
ccctccagct gctgaggttg tttaagttga aggagcgcta ccgcaacgtg ctggacacca 1920
tgttcgagct gctgccccgg atggccagcc tgggcctcac cctgctcatc ttttactact 1980
ccttcgccat cgtgggcatg gagttcttct gcgggatcgt cttccccaac tgctgcaaca 2040
cgagtacagt ggcagatgcc taccgctggc gcaaccacac cgtgggcaac aggaccgtgg 2100
tggaggaagg ctactattat ctcaataatt ttgacaacat cctcaacagc tttgtgaccc 2160
tgtttgagct cacagttgtc aacaactggt acatcatcat ggaaggcgtc acctctcaga 2220
cctcccactg gagccgcctc tacttcatga ccttttacat tgtgaccatg gtggtgatga 2280
cgatcattgt cgcctttatc ctcgaggcct tcgtcttccg aatgaactac agccgcaaga 2340
accaggactc ggaagttgat ggtggcatca cccttgagaa ggaaatctcc aaagaagagc 2400
tggttgccgt cctggagctc taccgggagg cacggggggc ctcctcggat gtcaccaggc 2460
tgctggagac cctctcccag atggagagat accagcaaca ttccatggtg tttctgggac 2520
ggcgatcaag gaccaagagc gacctgagcc tgaagatgta ccaggaggag atccaggagt 2580
ggtatgagga gcatgccagg gagcaagagc agcagcgaca actcagcagc agtgcagccc 2640
ccgccgccca gcagccccca ggcagccgcc agcgctccca gaccgttacc tagcccagcg 2700
cccgaaagcc gtctcttcta tgcaataaca caatagtatt actctactgc gatgtacgga 2760
actgcggtgt gtgtacacat actcacgtat atgcacatat ttatatacag gaagaaaaaa 2820
gacagacaag atggggcttg gtttataacc accttgccct gtcttcctta actccagaag 2880
ccagtttggt gaggggtggg ggtgcggcca ccaggtctga gctcttccta ctgtggaagg 2940
ctccagaagg cccttcacaa ggagacccct cacctggatc cagtcgactg cggggcttgc 3000
ccctcatgtg ggctggcctc catcggccac gtccaaagct gtcactgcta ctgcttcagg 3060
ctcacatccc cccgacctga tggcgtgccc gccccctctc cctgcggccc atgccacagg 3120
tttctgtgtt ttgctttagg gacagaacca cttaggaagg aaagaactcc cggtctccag 3180
ggtggtattt cagtgtctgt gataatgtca cgcaacacct cttcggggac c 3231

<210> 23
<211> 3160
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3968527CB1

<400> 23
atgacggaca acatcccgct gcagccggtg cgccagaaga agcggatgga cagcaggccc 60

cgcgccgggt gctgcgagtg gctgagatgc tgcggtggag gggaggccag gccccgcact 120
gtctggctgg ggcaccccga gaagagagac cagaggtatc ctcggaatgt catcaacaat 180
cagaagtaca atttcttcac ctttcttcct ggggtgctgt tcaaccagtt caaatacttt 240
ttcaacctct atttcttact tcttgcctgc tctcagtttg ttcccgaaat gagacttggt 300
gcactctata cctactgggt tcccctgggc ttcgtgctgg ccgtcactgt catccgtgag 360
gcggtggagg agatccgatg ctacgtgcgg gacaaggaag tcaactccca ggtctacagc 420
cggctcacag cacgaggcac agtgaaggtg aagagttcta acatccaagt tggagacctt 480
atcatcgttg aaaagaacca gcgggtccct gccgacatga tcttcctgag gacatcagaa 540
aaaaacgggt catgcttctt gcggacggat cagctggatg gggagacgga ctggaagctg 600
cggcttcccg tggcctgcac gcagaggctc cccacggccg ccgaccttct tcagattcga 660
tcgtatgtgt acgcagaaga gccaaatatt gacattcaca acttcgtggg aacttttacc 720
cgagaagaca gcgacccccc gatcagcgag agcctgagca tagagaacac gctgtgggct 780
ggcactgtgg tcgcatcagg tactgttgtg ggtgttgttc tttacactgg cagagaactc 840
cggagtgtca tgaatacctc aaatccccga agtaagatcg gcctgttcga cttggaagtg 900
aactgcctca ccaagatcct ctttggtgcc ctggtggtgg tctcgctggt catggttgcc 960
cttcagcact ttgcaggccg ttggtacctg cagatcatcc gcttcctcct cttgttttcc 1020
aacatcatcc ccattagttt gcgcgtgaac ctggacatgg gcaagatcgt gtacagctgg 1080
gtgattcgaa gggactcaaa aatccccggg accgtggttc gctccagcac gattcctgag 1140
cagctgggca ggatttcgta cttactcaca gacaagacag gcactcttac ccagaacgag 1200
atgattttca aacggctcca tctcggaaca gtagcctacg gcctcgactc aatggacgaa 1260
gtacaaagcc acattttcag catttacacc cagcaatccc aggacccacc ggctcagaag 1320
ggcccaacgc tcaccactaa ggtccggcgg accatgagca gccgcgtgca cgaagccgtg 1380
aaggccatcg cgctctgcca caacgtgact cccgtgtatg agtccaacgg tgtgactgat 1440
caggctgagg ccgagaagca gtacgaagac tcctgccgcg tataccaggc atccagcccc 1500
gatgaggtgg ccctggtaca gtggacggaa agtgtgggct taaccctggt gggccgagac 1560
cagtcttcca tgcagctgag gacccctggc gaccagatcc tgaacttcac catcctacag 1620
atcttccctt tcacctatga aagcaaacgt atgggcatca tcgtgcggga tgaatcaact 1680
ggagaaatta cgttttacat gaagggagca gatgtggtca tggctggcat tgtgcagtac 1740
aatgactggt tggaggaaga gtgtggcaac atggcccgag aagggctgcg ggtgctcgtg 1800
gtggcaaaga agtctcttgc agaggagcag tatcaggact ttgaagcccg ctacgtccag 1860
gccaagctga gtgtgcacga ccgctccctc aaagtggcca cggtgatcga gagcctggag 1920
atggagatgg aactgctgtg cctgacgggc gtggaggacc agctgcaggc agatgtgcgg 1980
cccacgctgg agaccctgag gaatgctggc atcaagggtt ggatgctgac aggggacaag 2040
ctggagacag ctacgtgcac agcgaagaat gcacatctgg tgaccagaaa ccaagacatc 2100
cacgtttttc ggctggtgac caaccgcggg gaggctcacc tcgagctgaa cgccttccgc 2160
aggaagcatg attgtgccct ggtcatctcg ggagactccc tggaggtttg cctcaagtac 2220
tatgagtacg agttcatgga gctggcctgc cagtgcccgg ccgtagtctg ctgccgatgt 2280
gcccccaccc agaaggccca gatcgtgcgc ctgcttcagg agcgcacggg caagctcacc 2340
tgtgcagtag gggacggagg caatgacgtc agcatgattc aggaatctga ctgcggcgtg 2400
ggagtggaag gaaaggaagg aaaacaggct tcgttggctg cagacttctc catcactcaa 2460
tttaagcatc ttggccggtt gcttatggtg catggccgga acagctacaa gcggtcagcc 2520
gccctcagcc agttcgtgat tcacaggagc ctctgtatca gcaccatgca ggctgtcttt 2580
tcctccgtgt tttactttgc ctccgtccct ctctatcaag gattcctcat cattgggtac 2640
tccacaattt acaccatgtt tcctgtgttt tctctggtcc tggacaaaga tgtcaaatcg 2700
gaagttgcca tgctgtatcc tgagctctac aaggatcttc tcaagggacg gccgttgtcc 2760
tacaagacat tcttaatatg ggttttgatt agcatctatc aagggagcac catcatgtac 2820
ggggcgctgc tgctgtttga gtcggagttc gtgcacatcg tggccatctc cttcacctcg 2880
ctgatcctca ccgagctgct catggtggcg ctgaccatcc agacctggca ctggctcatg 2940
acagtggcgg agctgctcag cctggcctgc tacatcgcct ccctggtgtt cttacacgag 3000
ttcatcgatg tgtacttcat cgccaccttg tcattcttgt ggaaagtctc cgtcatcact 3060
ctggtcagct gcctcccccc ctatgtcctc aagtacctgc gaagacggtt ctctcccccc 3120
agctactcaa agctcacatc ataggccgtg cgttcgctgg 3160

<210> 24
<211> 2848
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472732CB1

<400> 24

cttaacactg aacccattac ttttccaaga ccagaaaaaa atattacatg aacaggaact 60
acttctcctt cagataagaa ttcaagcttt gacattgtaa accacagacg aattggagct 120
cggcattgaa aggaggtgtt ctgcaatgat ttttttttctt gtttagagaa gtttacttct 180
acaagaagaa atctgaaaaa tgacaggagc aaagaggaaa aagaaaagca tgctttggag 240
caagatgcat acccccccagt gtgaagacat tatacagtgg tgtagaaggc gactgcccat 300
tttggattgg gcaccacatt acaatctgaa agaaaacttg cttccagaca ctgtgtctgg 360
gataatgttg gcagttcaac aggtgaccca aggattggcc tttgctgttc tctcatctgt 420
gcacccagtg tttggtttat atgggtctct gtttcctgcc ataatttatg ccatatttgg 480
aatgggacat catgttgcca caggcacctt tgccttgaca tccttaatat cagccaacgc 540
cgtggaacgg attgtccctc agaacatgca gaatctcacc acacagagta acacaagcgt 600
gctgggctta tccgactttg aaatgcaaag gatccacgtt gctgcagcag tttccttctt 660
gggaggtgtg attcaggtgg ccatgtttgt gctgcaactg ggcagtgcca catttgtggt 720
cacagagcct gtgatcagcg caatgacaac tggggctgcc acccatgtgg tgacttcaca 780
agtcaaatat ctcttgggaa tgaaaatgcc atatatatcc ggaccacttg gattctttta 840
tatttatgca tatgtttttg aaaacatcaa gtctgtgcga ctggaagcat tgcttttatc 900
cttgctgagc attgtggtcc ttgttcttgt taaagagctg aatgaacagt ttaaaaggaa 960
aattaaagtt gttcttcctg tagatttagt tttggctcca aacacatcgc cactccatca 1020
ccactacgac tgtctctttg ccaactttct tgagccaccc tgggaggatg gacttccaga 1080
aggtgccttc aaccaggcag aaggacattt gcgcaggaac ataattccct cacctagagc 1140
tcccccgatg aacatcctct ctgcggtgat cactgaagct ttcggagtgg cacttgtagg 1200
ctatgtggcc tcactggctc ttgctcaagg atctgccaaa aaattcaaat attcaattga 1260
tgacaaccag gaatttttgg cccatggcct cagcaatata gtttcttcat ttttcttctg 1320
cataccaagt gctgctgcca tgggaaggac ggctggcctg tacagcacag gagcgaagac 1380
acaggtggct tgtctaatat cttgcatttt cgtccttata gtcatctatg caataggacc 1440
tttgctttac tggctgccca tgtgtgtcct tgcaagcatt attgttgtgg gactgaaggg 1500
aatgctaata cagttccgag atttaaaaaa atattggaat gtggataaaa tcgattgggg 1560
aatatgggtc agtacatatg tatttacaat atgctttgct gccaatgtgg gactgctgtt 1620
tggtgttgtt tgtaccatag ctatagtgat aggacgcttc ccaagagcaa tgactgtaag 1680
tataaaaaat atgaaagaaa tggaatttaa agtgaagaca gaaatggaca gtgaaaccct 1740
gcagcaggtg aaaattatct caataaacaa cccgcttgtt ttcctgaatg caaaaaaatt 1800
ttatactgat ttaatgaaca tgatccaaaa ggaaaatgcc tgtaatcagc cacttgatga 1860
tatcagcaag tgtgaacaaa acacattgct taattcccta tccaatggca actgcaatga 1920
agaagcttca cagtcctgcc ctaatgagaa gtgttattta atcctggatt gcagtggatt 1980
tacctttttt gactattctg gagtctccat gcttgttgag gtttacatgg actgtaaagg 2040
caggagtgtg gatgtattgt tagcccattg tacagcttcc ttgataaaag caatgacgta 2100
ttatggaaac ctagactcag agaaaccaat ttttttgaa tcggtatctg ctgcaataag 2160
tcatatccat tcaaataaga atttgagcaa actcagtgac cacagtgaag tctgagaccc 2220
ttttgtcaca gtacagctct tgtctttacc aactgcctga agaggccata tgctggcatt 2280
ttgcacaact ttttggttgt ttagatccta cagatgacct ctgctacaat aagtacgatg 2340
tgacttagta actgcatagc agttggaaag aactgccaac tttttttttct cattttgtt 2400
agtaagaaga ttcgcttagt tattttatgt aaaaatcagt atgtgtttag ttttagtgta 2460
ctgaagggta aacatggttt tattttattt taccatatta ttttgtgttg ttttatttct 2520
attgtgctgt aagttgatgt ttaaaattga gaaatacttt tgtcataggt aatttggaac 2580
atttacaagc catttgtaaa attttaagat aatctgtaac taatacataa aaacaactta 2640
gcaaatgtgc cattttcaca caacttctct ctgtataggc ctctgaaata tcaataaggc 2700
taaatattac tttacacagt aagatgtgaa attcacaaaa agtaaaccaa actaaacgaa 2760
tgaaaaactg gaaataattc gtttccatat ctttccatac gtccatttct gaagtattca 2820
ggaatgtttt cataatcgaa agaaacgg 2848

<210> 25
<211> 3727
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476938CB1

<400> 25
atggttatgg aggctgggga gtccaagggc atagtgctgt catctggcaa gggccttcat 60
gctgcatcat tcatggtgga aggtgaaaac gtaagagaag ggattggctc agaaatgggc 120
acctgcccca agtggaccaa tgtttctcat tgcaaaatgg gaataatgcc agttttggtt 180
aagggcttcg tgctgagcgg aagccggaag caaaagcggg tcctgctagc cccgcggctc 240

cgaactcggt ggtcctggaa gctccgcagg atgggggaga agatggcgga agaggagagg 300
ttccccaata caactcatga gggtttcaat gtcaccctcc acaccaccct ggttgtcacg 360
acgaaactgg tgctcccgac ccctggcaag cccatcctcc ccgtgcagac aggggagcag 420
gcccagcaag aggagcagtc cagcggcatg accattttct tcagcctcct tgtcctagct 480
atctgcatca tattggtgca tttactgatc cgatacagat tacatttctt gccagagagt 540
gttgctgttg tttctttagg tattctcatg ggagcagtta taaaaattat agagtttaaa 600
aaactggcga attggaagga agaagaaatg tttcgtccaa acatgttttt cctcctcctg 660
cttcccccta ttatctttga gtctggatat tcattacaca agggtaactt ctttcaaaat 720
attggttcca tcaccctgtt tgctgttttt gggacggcaa tctccgcttt tgtagtaggt 780
ggaggaattt attttctggg tcaggctgat gtaatctcta aactcaacat gacagacagt 840
tttgcgtttg gctccctaat atctgctgtc gatccagtgg ccactattgc cattttcaat 900
gcacttcatg tggaccccgt gctcaacatg ctggtctttg gagaaagtat tctcaacgat 960
gcagtctcca ttgttctgac caacacagct gaaggtttaa caagaaaaaa tatgtcagat 1020
gtcagtgggt ggcaaacatt tttacaagcc cttgactact tcctcaaaat gttctttggc 1080
tctgcagcgc tcggcactct cactggctta atttctgcat tagtgctgaa gcatattgac 1140
ttgaggaaaa cgccttcctt ggagtttggc atgatgatca tttttgctta tctgccttat 1200
gggcttgcag aaggaatctc actctcaggc atcatggcca tcctgttctc aggcatcgtg 1260
atgtcccact acacgcacca taacctctcc ccagtcaccc agatcctcat gcagcagacc 1320
ctccgcaccg tggccttctt atgtgaaaca tgtgtgtttg catttcttgg cctgtccatt 1380
tttagttttc ctcacaagtt tgaaatttcc tttgtcatct ggtgcatagt gcttgtacta 1440
tttggcagag cggtaaacat tttccctctt tcctacctcc tgaatttctt ccgggatcat 1500
aaaatcacac cgaagatgat gttcatcatg tggtttagtg gcctgcgggg agccatcccc 1560
tatgccctga gcctacacct ggacctggag cccatggaga agcggcagct catcggcacc 1620
accaccatcg tcatcgtgct cttcaccatc ctgctgctgg gcggcagcac catgcccctc 1680
attcgcctca tggacatcga ggacgccaag gcacaccgca ggaacaagaa ggacgtcaac 1740
ctcagcaaga ctgagaagat gggcaacact gtggagtcgg agcacctgtc ggagctcacg 1800
gaggaggagt acgaggccca ctacatcagg cggcaggacc ttaagggctt cgtgtggctg 1860
gacgccaagt acctgaaccc cttcttcact cggaggctga cgcaggagga cctgcaccac 1920
gggcgcatcc agatgaaaac tctcaccaac aagtggtacg aggaggtacg ccagggcccc 1980
tccggctccg aggacgacga gcaggagctg ctctgacgcc aggtgccaag gcttcaggca 2040
ggcaggccca ggatgggcgt ttgctgcgca cagacactca gcaggggcct cgcagagatg 2100
cgtgcatcca gcagcccctt caagacataa gagggcgggg cgaggtactg gctgcagagt 2160
cgccttagtc cagaacctga caggcctctg gagccaggcg acttcttggg aaactgtcat 2220
ctcccgactc ctccctgagc cagcctccgc tcagtgtggc tcctcagccc acagagggga 2280
gggagcatgg ggccaggtgc cagtcatctg tgaagctagg gcgcctaccc ccccacccgg 2340
aggacccctg cggcccctg cctagaggag caccatctac agttgtgcca ttccccagcc 2400
actgccttca tgctgccccc gccggactgg cagagccagg ggtcagccac ctgcctttga 2460
gtcatcaaga tgcctctgca gccacaattc tgacctaagt ggcagggccc agaaatcctg 2520
aaaacctccc gctgcctttt gtgatacttc ctgtgctccc tcagagagaa acggagtgac 2580
cttttgtcct ttacctgatt ggcacttcgc agtctatctc cctgggtagc agacggctgc 2640
tgcccttctc tgggcatgtt ctgaatgttt acactggtac cttctggtat cttctttaga 2700
gccccctgca agctgcaact ctaggctttt atcttgcggg gtcagagcgc cctctagagg 2760
gaaaagctag aggcacaggg tttctgccgg cccacaactg ctgtcttgat ttgcatttta 2820
cagcaaagtg ctgagagcct ctagtcgcct cctgccatct gatctccctc cccaccattc 2880
ccgtactcag ttgttctttt gtctaatcgg aggccactgt gctgaggccc tgcagtgtct 2940
gctcactgct gccatcttcg ctgctagtca gggttccatc ctctttcccc tctcccagtt 3000
ccctaccacg ttggatccca ttcgtcaccc atgctagggt ccccaaagca ctggggcagg 3060
ggccagagca gcagcaccca gtgctccctc ctctactctg acctggggcc ccagcatcct 3120
ggagcacacg ctccacgcac acacacccca gccctgtccc aggggcctgg ccccctcagc 3180
catctcaggg tgaggagctg ccagtcatgt ccagatggaa tgactcccat cctctcctca 3240
tctccccttt gacgagcctc aaactgctca gctcatcaaa gagccattgc caacttccgt 3300
atgtggttct gggtcccagg gagccttgga acctggcacc ctggggtggt ttaattcatc 3360
attaagaagc attcctgctt ctcaagggac acagtggcct gcatgggcca gcatggaccc 3420
tgggctgatc atgtgcattc ctgcttctct ggggacacag tgggcccaca tgggccagca 3480
tggaccctgg gctagagcaa gcacatctcc atctcttcca cctcaggcag tgtggctcca 3540
gatgtcagga gggactgacc tcaggacctt ccaggttcct ctgtgccagg aatgagaggc 3600
caggcccgat cctaccacct cgccttgacc ctgaagtcag agcaggccag ccaagcagga 3660
agcacactgt ttacttttg catgaaaagt aaatgtgtac ttgatagagc taaaatatga 3720
tcttttt 3727

<210> 26
<211> 2571
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8128531CB1

<400> 26
ttaaagctgg acagaatttt taaaagcaat gaagccagtt ccttggatat atccacgggc 60
tttgctttga gaaggaactg agtaggcagt gagaagagtc gagtgacgcc tggcccgtga 120
gtgcctcaac aactgagatg aacgtcgact cgcttgcagg caagttgtca gggaggaagc 180
cacagcccaa ggaggtcgtc acttgccggg aaggtggctc gggccaggct gcactcaaaa 240
cccgtgctct gtccacactg ctacggggcc agagccaagg aagcttccac ttcttccccc 300
agacagcccc aacagcggct accccaagga gccagcagcc ttgtgtcctg ggatccccag 360
cccctgcaga atgacccacc aggatctgag catcacagcc aaactcatca atggaggtgt 420
agcagggctc gtgggggtga cctgcgtgtt ccccatcgac ttggccaaga ctcgcctgca 480
gaaccagcat gggaaagcca tgtacaaagg aatgatcgac tgcctgatga agacggctcg 540
ggcggagggc ttcttcggca tgtaccgagg ggctgcagtg aacctcactc tggtcactcc 600
agagaaggcc atcaagctgg cggccaacga ctttttccgg cggctgctca tggaagatgg 660
gatgcagcgg aacctgaaga tggagatgct tgccgggtgt ggggctggga tgtgccaggt 720
cgtggtgacc tgtcccatgg aaatgctcaa gattcagctg caggatgctg gacgcctggc 780
cgtccatcat cagggctcgg cctcagcacc ctccacctcc aggtcctaca caactggttc 840
ggcttccacc cacaggcgcc cctctgccac cctcattgcc tgggagctgc tccgcactca 900
gggcctggct gggctctaca ggggcctggg tgccactctc ctcagagaca ttcctttctc 960
catcatctac ttcccactgt ttgccaacct taacaacctg gggttcaacg agctcgccgg 1020
taaggcgtcc tttgcacatt ccttcgtgtc aggctgtgtg gcaggttcca tagctgcggt 1080
cgcagtgacg cctctagatg ttctgaaaac tcgaatccaa accctcaaga aaggcctggg 1140
cgaggacatg tacagtggga tcaccgactg tgccaggaaa ctctggattc aggagggacc 1200
atctgccttc atgaaaggcg ctggctgccg ggcactggtc atagcacctc tctttgggat 1260
tgctcaaggg gtctatttta ttgggattgg agagcgcatc ttaaagtgtt ttgactagac 1320
agagctggag gtcaagtccc tgcgcttgcc gccctctctc tagctgtttc acttagccta 1380
gaggggcaa gggcaggtgg ggccactctg gcctgcctgg tcctctgcgt tgtagtgcta 1440
cctcaatctc gggagaaaca gccctatatt ctaacaagtt gagcacagcc ttcttccct 1500
tcgtgtctac actcgttttc ctttgtgggc acagctacca ggggctttg gaagccccta 1560
accacctact tttcaacaaa aatggtactt tcgttgtatt aattgcagga ccttaacagg 1620
tagtcacaat agaagggttg tttctgtatt ttaacatttc tatttcacag tcaaactcgg 1680
cattcttcag tcagcttgag gatttagcat tgttaatctt ggactccata acttatgagt 1740
cctagcactg attttgagga aaaggaggat cagaagttca agggaccgtg aaagccctca 1800
gagtcagcac ctagtttgag accaagcacc ctttcgaatc cctggatggc tgaggggct 1860
gaggccggct ctgactgggc agctcagccc ctccccaga gcccagggtc ttgcacaccc 1920
ctccctgtaa ccaaggaaca ctctgaaata aaggtgaatg gctaaaatct catctgttca 1980
tcagtgggta cagcagatag gctgcagtga atgctatcac catctacttt tctacgtcca 2040
tttcaaaacc aaacattaaa aagggcatag aagcagaccc ccgtcactct tcaaactgtt 2100
acttgtgggg gtggaggaac acagccatag ggaaatatct gcttgttagt gacactgggt 2160
tttaagcctt gattctatcc cttcataagt gaatcgtctt gaggagctga gtttgctgtg 2220
agagccctcc tcacgcacct cgattcctcc cccaaaggct gctacaggag agataatgtc 2280
acagcagcag ggccaagtcc taagaaaatc agcacctgct gcaggagctg gtgtttacaa 2340
tagtcccatc tactgtgaaa cctgggctaa caaggaagag gatggtgcta acatggtcag 2400
ccctgggggc ctcactctct gttatgagaa ctgcatttga gtatgggccc tggagacaga 2460
cctcagttca agtcccagct ccaccatgta ctagctgcaa ggccctgggc agctcttagt 2520
cgtcacctac ggaaaaataa aacatgggac agggaaggaa gaacagggcc t 2571

<210> 27
<211> 1660
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476757CB1

<400> 27
ggagttctgg gctgtagtgc gctatgccga tcgggtgtcc gcactaagtt cggcatcaat 60
atggtgacct cccgggagcg ggggaccacc aggtcgcctc tcattctgta caaaggtgtg 120

ctatggcatt ggtggggtcc ccaaccagat agcctccagc gccacagcct tttacctgca 180
gcttttcctg cttgatatag cacagatccc tgccgcccag gtgtcacttg ttctgtttgg 240
gggaaaggtg tctggggcgg ctgctgaccc tgtggctggg ttcttcatca acaggagcca 300
gaggacaggg tctggacggc tcatgccttg ggtgctgggc tgcacccccc tcatcgccct 360
ggcctacttc ttcctgtggt tcctgccccc cttcaccagc ctgcgaggcc tctggtacac 420
gactttctac tgcctgttcc aggccctggc cacgttcttc caggtgccct acacagcgct 480
caccatgctg ctgactccct gcccaaggga gcgggactcg gccaccgcca taccggatga 540
ctgtggagat ggcgggaaca ctgatggggg ccactgtcca cgggctcatc gtgtccggcg 600
cccacagacc ccacaggtgc gaggccactg cgaccccggg gccagtcact gtctccccga 660
atgcagccat ctctactgca ttgcggctgc cgtggttgta gtgacttacc ccgtgtgcat 720
cagtttactg tgcctagggg tgaaggagcg gccaggtttt gcttttgaac tctgcgaagc 780
caaggtgaca cgcttctgcg ttgcagaccc ctctgcccca gcctcaggcc caggcttgag 840
tttcctggct gggctgagcc tcactacccg gcacccaccc tacctgaagc tggtgatctc 900
cttcctgttc atctctgctg ctgttcaggt ggagcagagc tacctggtcc tgttctgtac 960
acatgcctcc cagctacacg accacgtcca gggcctggtc tcagccgtgc tgagcacccc 1020
gctgtgggag tgggttctcc agcgctttgg gaagaagacg tcagcctttg ggatctttgc 1080
gatggtgccc tttgcgatct tgctggctgc tgtgcccaca gcacctgtgg catatgtcgt 1140
ggcctttgta tctggcgtga gcattgctgt gtccttgctg ctaccctggt ccatgctgcc 1200
agacgtggtg gatgactttc agctgcagca ccgtcacggg ccaggcctgg agaccatctt 1260
ctactcctcc tacgtcttct tcaccaagct gtctggcgca tgtgccctgg gcatctccac 1320
cctcagtctg gagttctcgg ggtataaggc aggggtctgc aagcaagcag aggaggtggt 1380
ggtcaccctc aaagtcctca ttggcgccgt gcccacctgc atgatccttg ctgggctctg 1440
catcctcatg gtcggctcca ctccaaagac acccagtcgg gacgcctcca gccggctgag 1500
ccttcggaga cgtgcacaag cacccaatgt tcacacaagt aaggtccacg agcatgcaca 1560
tatcatgcag gcccacgcgg gacaggcagt gggtggcctt gtcatcagcc actccctgct 1620
gagggtgacg gcctcgggct ctgcagcaga gagatactga 1660

<210> 28
<211> 2743
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 266243CB1

<400> 28
atggcggcgg ccgcggtggg cgcgggccac ggcgcggggg gcccgggcgc agcgagcagc 60
agtggtgggg cgcgcgaggg cgcgcgggtg gcggcgctgt gcctgctgtg gtacgcgctg 120
agcgcgggcg gcaacgtggt caacaaggtg atcctgagcg ccttcccgtt cccggtgacc 180
gtgtcgctgt gccacatcct ggctctgtgc gctgggctcc cgccgctgct gcgcgcctgg 240
cgcgtgcccc ccgcgccgcc cgtctcgggc cccggaccca gtccgcatcc gtcgtccggc 300
ccgctgctgc cgccgcgctt ctacccgcgc tacgtgctac cgctcgcctt cggcaagtac 360
ttcgcgtccg tgtcagcgca cgtcagcatc tggaaggtgc ccgtgtccta tgcacacacc 420
gtcaaggcca ccatgcccat ctgggtggtc ctcctgtccc ggatcattat gaaggagaag 480
cagagcacca aggtatactt gtcactcatc cccatcatca gcggtgtcct gctggccacc 540
gtcaccgagt tgtcttttga catgtgggga ctcgtcagcg ccctcgccgc cacgctgtgc 600
ttctcgcttc agaacatttt ctccaaaaag gtcttgcgag attcacggat ccaccatctc 660
cggctgctca acatcctggg ctgccacgcc gtcttcttta tgatccccac ctgggttctg 720
gtggacctct cggctttcct ggtcagcagc gacttgacct acgtctacca gtggccctgg 780
acgctcctgc tcctggctgt cagcggcttc tgtaactttg cccagaatgt tatcgccttc 840
agcatcctca acctcgttag cccgctgagc tactcggtcg ccaatgccac caaaagaatc 900
atggtcatca cggtgtccct gatcatgctg cgcaacccag tcaccagcac caacgtcctg 960
ggcatgatga ccgccatcct gggggtcttc ctctataaca agaccaagta cgatgcaaac 1020
cagcaagcca ggaagcacct cctccccgtc accacagcag acctgagcag caaggagcgt 1080
caccggagcc cactggagaa gccccacaac ggcctcctct tcccccagca cggggactat 1140
cagtacggcc gcaacaacat cttaacagac cacttccaat acagccggca gagctaccca 1200
aactcgtaca gtttgaaccg ctatgatgtg tagagtccaa aggacaggac cagactgttg 1260
gtgactcctt ccccggcccc cacagcagta tcagaaactt ctgacaatca gtgaatgtac 1320
aacccagccg aggggacggt gcataactct ccatcagaag ccctggggtt cctggccccc 1380
cgtgagccgc aggaggatgc gttgcctgca gtgcagacgg ccgtgagctc tgggcaaacc 1440
taaacagaga ccagtgtctc atgctctttc ttcctggagt ctgtcatctg agggccgtgt 1500
ccctgcggag atcttggcca cgttgtacct ttccatgtgg aattattccc caagcagtgt 1560

agctcagagc acttgtgtct gcattccaga taacattcag gacctgtgtg aaaagctggg 1620
gtcactgtgg ctgtagacca tgaactggca gtgggggtgt ccagggcggt gcttgagaac 1680
gtcagactgg ctagtttaat tccctggcgc agatacgcat aggaccaaca gggtcaccaa 1740
gcagacaggg agcccgcgag aatcattcaa aacatcccca gccacagaga tggatccagt 1800
ttcctggtca tcccccttagc agttcacaag ttcctggcaa atgttccaaa gcaaaaagcg 1860
attgcaatta gcatccagtt cctgcagcct ggtgctctgc cctgcacgtc agggttggca 1920
tccacccaga tccagatgga agggaaactt ctctcttctc ctttgcctcc tcttccctca 1980
ccagagcagg gcgcttctct tggggtggtg agaaggatct tcgagaaatc gtgttcagta 2040
tttcaagctc tatttctgtg gcacatgtct tttgagaggc atcttcacct cttctgtgat 2100
gacttggtat gttgtttggt agagagatct tgattttcgg aggatcttgc atttttctag 2160
ggaatatttt gtagttgtgt gtgtgtgttt ttgccttggt ccccattatg ggatgcatta 2220
ggactggcct atgcatcgaa aatctttttg tttgtaaacg tttaaaaaca aagttccccg 2280
gccaggcaca gtggctcaca cctgtagtcc cggcactttg ggaggccaag atgggcggat 2340
cacgaggtca ggagttcgag accagcctgg ccaacatggt gaggccccgt ctctactagg 2400
agtacagaaa ttagccgggc atggtgtcgc gtgcctgtgg tcccagctcc tcgggctgct 2460
gaggcaggcg aattgcttga acctgggagt gcagtgatgc gacctcggct gactgcaacc 2520
tctacctccc gggttcaaac aattgtcttg tctcagcctc ccaagtagct gggattacag 2580
gtgtgcacca ccatgcctgg ctaattttta gtagagatgg ggtttcacca tgttggcctg 2640
gatggtctcg aactcctgac ctcagatgat ccacctgcct tggcctccca aagtgctggg 2700
attacaggca tgagccacca cacaaccgac cttggccagc aca 2743

<210> 29
<211> 3239
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6585710CB1

<400> 29
ggcttaaagt aggggcggcc agcacatggt ccattatttc acagccatcg gctacccctg 60
tcctcgctac agcaatcctg ctgacttcta tgtggacctg accagcattg acaggcgcag 120
cagagagcag gaattggcca ccagggagaa ggctcagtca ctcgcagccc tgtttctaga 180
aaaagtgcgt gacttagatg actttctatg gaaagcagag acgaaggatc ttgacgagga 240
cacctgtgtg gaaagcagcg tgaccccact agacaccaac tgcctcccga gtcctacgaa 300
gatgcctggg gcggtgcagc agtttacgac gctgatccgt cgtcagattt ccaacgactt 360
ccgagacctg cccaccctcc tcatccatgg ggcggaggcc tgtctgatgt caatgaccat 420
cggcttcctc tattttggcc atgggagcat ccagctctcc ttcatggata cagccgccct 480
cttgttcatg atcggtgctc tcatcccttt caacgtcatt ctggatgtca tctccaaatg 540
ttactcagag agggcaatgc tttactatga actggaagac gggctgtaca ccactggtcc 600
atatttcttt gccaagatcc tcggggagct tccggagcac tgtgcctaca tcatcatcta 660
cgggatgccc acctactggc tggccaacct gaggccaggc ctccagccct tcctgctgca 720
cttcctgctg gtgtggctgg tggtcttctg ttgcaggatt atggccctgg ccgccgcggc 780
cctgctcccc accttccaca tggcctcctt cttcagcaat gccctctaca actccttcta 840
cctcgccggg ggcttcatga taaacttgag cagcctgtgg acagtgcccg cgtggatttc 900
caaagtgtcc ttcctgcggt ggtgttttga agggctgatg aagattcagt tcagcagaag 960
aacttataaa atgcctctcg ggaacctcac catcgcggtc tcaggagata aaatcctcag 1020
tgccatggag ctggactcgt accctctcta cgccatctac ctcatcgtca ttggcctcag 1080
cggtggcttc atggtcctgt actacgtgtc cttaaggttc atcaaacaga aaccaagtca 1140
agactggtga ttcacgccag acgtctgccc gctggtgggg gacctgagca gacccttcaa 1200
ctgcactccc tcctcaggag ccccttcctg gggacagtga ggacaatgac cctacagatg 1260
ctcagctaca tccggcccag ggtgctgcag tggcacagac cagccacagg atggcagtag 1320
aataaagaca gtcgaaaggg atttctgctc actggcagga gactgcgatg actgggagaa 1380
aacctgcact cggtggcacc tacaacgttg ctaatttatt tccttttgat atgcatttat 1440
ataggcaact cgatatagga tgggagcaaa ctaggaatga attgggtagc tagactgtgc 1500
aggaattgtt ggaacctgga gggaacaata acagtagcta gcagatttgg cttcatcttc 1560
caggggcccc acactccgtg gtgagccacc atcaatacag aaagtgacct aagatgtacc 1620
agcaagatgc catcccttct ttttgtgtgg ggtcatgggc tccaaaagcc aacgtgaaca 1680
attaaaaatg tattgagcat ctactctgta gcaggtcctg tgaaaacact ttaggtggac 1740
aatcccttga ggtaagtggt atcccatttt ataggtgtga aaactgaagc aaaaattcat 1800
tttcctaagg gcacatggat acttgtggtg gagtcatatg gggatcagaa aagcctttga 1860
ggccttggag ttagagggca gaaggcaagg cctgagccgc tgtaagccct taggagttta 1920

ggaaggctcc agaagacaaa tggggtctgt agaggctgtt aactcagcca ggcttcttag 1980
agttgcattt cactaactga tatggtttgg ctctgcgtcc tcacccaaat ctcaccttga 2040
attgtaataa tccccaagtg tcaagggcgg gaccagatgg agataattga atcataggg 2100
tggtttctct gatgctgttc tcctgagagt gagtgagttc tgatgagatc cgacggtttt 2160
ataaggggct tccctcttcg ctcggctctc attctctctc ccgctaccct gtgaagagga 2220
gccttccacc acgactgcaa gtttcctgag gctgccccag ccgtgctgaa ctgtgagtca 2280
gttaaacctc ttttctttat aaattaccca gtcttgggta tttcttcata gcagtgtgag 2340
agcagatgaa tacactggcc ctgcctgggt ttcagaacca gccttgaacc tttcacagtg 2400
gccagaggat ggggaggcag aggcccaggt tgcacacttc ttgcctgagt gttggggact 2460
atctgaccca aaacaggtgc acagagggca ggagaggatg ttcccaaagg aaaattagag 2520
tttagaatca aaagaagggg aaggtgcgtg tttgggaggt aaatagcaaa tactcttcat 2580
aggttcacta gagtcttgct actccaagta cgatccctgg gccagcagaa tgggcacagc 2640
tggagctgat tggataggtc ccatgagcct caggccccac ccaggctcaa tgagtcagag 2700
tctgcgtctt aagaagaccc cctggtgatc tgtgcacatt caagcatgct gccgttttcc 2760
aaagcacttg caacactcag gatgcttgca cggtcatgtt gccaccatcc aacctgcaga 2820
ccccattctt gagattgact gggagttcct atcatgtcct ccatagcaag gggatctaga 2880
ccagaatcaa gccttggatc tagttctcaa gtctctttgt ctctttcagt ttaggaacag 2940
tttgtcaact ttccttcact ttgtgacctt gatacttgag tttgaaggct gtctctcaat 3000
ttgtgttctg cccagtgcat cctatcagaa ggcatgtgat ttcaacttct cccataccaa 3060
caacgttcac tttgatcact tgattaaagg ggtgtctgct aggcttctcc acagccaagt 3120
tactattttc cctcccttta taattaataa gcattttgta agtgggtact ttgaaactat 3180
gtaaattgta aactttccat ttatgcattt taaaattttg attgatgtaa aaaaaaaaa 3239

<210> 30
<211> 1615
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483599CB1

<400> 30
atggacaaat tcctagacac atacaatcta ccaagattga accaggaaga aatccaaaac 60
ctgaagagac caataacaag taatgagatc aaagcaataa taaaaagtct ccagatgtca 120
ttgcttggaa gggactacaa cagtgagctg aactccttgg acaacggacc tcagtcaccc 180
tcagagagca gcagtagcat tacttcagag aatgtccatc ctgctggaga agctggacta 240
tcgatgatgc aaactttgat ccacttgttg aaatgcaaca ttggcacagg gctcctgggg 300
cttcccctgg ccataaagaa tgccggcttg ttggtcggtc ctgtcagcct tctggccatc 360
ggggtcctca ccgtgcactg catggtcatc ctgttgaact gtgctcaaca cctcagccag 420
cctagactgc agaagacttt tgtgaactat ggagaggcca cgatgtacgg ccttgaaacc 480
tgcccgaaca cctggctgag ggcccatgca gtgtggggaa ggtacactgt cagcttctta 540
ttagtcatca cccagctggg cttctgcagt gtttatttta tgtttatggc agacaattta 600
caacagatgg tggaaaaagc ccacgtgacc tccaacatct gccagcccag ggagattctg 660
acgctgaccc ccatcctgga cattcgtttc tacatgctga taatcctgcc cttcctgatc 720
ctgttggtgt ttatccagaa cctcaaggtg ctgtccgtct tctcgacatt ggccaacatc 780
accacccttg ggagcatggc tctgatcttt gagtatatca tggaggggat tccatatccc 840
agcaacctac ccttgatggc aaactggaag accttcttgc tgttctttgg tacagccatc 900
ttcacatttg aaggcgtcgg tatggttctg cctctcaaaa accagatgaa gcatccacag 960
cagttttctt ttgttctgta cttggggatg tccattgtca tcatcctcta tatcttactg 1020
gggacactgg gctacatgaa gtttgggtca gacacccagg ccagcatcac cctcaacttg 1080
cccaattgct ggtatgtcct gcccacctca ggtgagatag ggagagacac tggaactgtt 1140
ctggttgtca tagcagagag cacagcaaag ctgagccatg aagctggtaa tccatcactg 1200
gaagtgacat atgtctctcc tgctcacact gcatcagtca aagcaagcca catggccgca 1260
cctcactcca agggggcagg gaagtgcaat tctgccatgt gcctggaagt atttggtgaa 1320
cagcacaaat aactgctgtg ctacctgatg cagtgaccgt ggggattaat tggattaata 1380
catagataat gcttagaaaa gtgcttagca cgtggtaagc attcatgagc gttagctatt 1440
atcattgtta tgcatcccca cagccttcat ttttccaagg tgagtaggat gatggtgcat 1500
ttatttccca caaatccaga gctgtagaat gagaaaaatg taaccatccc cacccacctt 1560
gctgtgttat gataatgact agatgagaca ataaatgtgg agtttctttg aaaaa 1615

<210> 31
<211> 1245

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2507246CB1

<400> 31
atggcgacgg gcggccagca gaaggagaac acgctgcttc acctcttcgc cggcgggtgt 60
ggaggcacag ttggtgctat tttcacttgt ccactagaag tcattaagac acggttgcag 120
tcttcaagat tagctctccg gacagtctac tatcctcagg ttcatctggg gaccattagt 180
ggagctggaa tggtgagacc aacatccgtg acacctggac tctttcaggt tctgaagtcg 240
atcttggaga aagagggacc aaagtcactt tttagaggct tgggtccaaa tttggttgga 300
gttgcaccat caagggctgt atactttgca tgttactcca aagccaaaga gcaatttaat 360
ggcattttcg tgcctaacag caatattgtg catattttct cagctggctc tgcagctttt 420
atcacaaatt ccttaatgaa tcctatatgg atggttaaaa cccgaatgca gctagaacag 480
aaagtgaggg gctctaagca gatgaataca ctccagtgtg ctcgttacgt ttaccagacc 540
gaaggcattc gtggcttcta tagaggatta actgcctcgt atgctggaat ttccgaaact 600
ataatctgct ttgctattta tgaaagttta aagaagtatc tgaaagaagc tccattagcc 660
tcttctgcaa atgggactga gaaaaattcc acaagttttt ttggacttat ggcagctgct 720
gctctttcta agggctgtgc ctcctgcatt gcttatccac acgaagtcat aaggacgagg 780
ctccgggaag agggcaccaa gtacaagtct tttgtccaga cggcgcgcct ggtgttccgg 840
gaagaaggct accttgcctt ttatagagga ctgtttgccc agcttatccg gcagatccca 900
aatactgcca ttgtgttgtc tacttatgag ttaattgtgt acctgttaga agaccgtact 960
cagtaacagg ccggaaaatt gtgctctaga agaataaaac tgaaaaactc tagagaattt 1020
tttttcccca ttgatgttta gaaagtttga gactgaaaca ggaaaggcca taaaatatct 1080
ggttcatatc acctgttgga catttccttt tggattcatg ctttctggaa ggtttaaatt 1140
cattaacgtt aatagttaat tataactttt tttttaactt aagaggattc agggttaagc 1200
accaactaaa ttaaatcatg ctatttaatt taagtataaa aaaaa 1245

<210> 32
<211> 4169
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3033505CB1

<400> 32
gcacggctca ctgacacgca gctttggtta aagagcgggc gcacaggagg ggaggagacc 60
gcgcgcggga cggggaggaa tggcctgtcc gcgttaaacc atcacaagcc atggttgcgg 120
aagggccacg cgtcccccag taggagaatg actccgattc gtgaccctca gcgccggtgc 180
atgtcgatat atttattgag tgtctactgt gtgccaggca ctatatctat gtgcatagaa 240
aaaccctgga aggccataca acaatatata tagagtgatc gtctctgctt gctgagctaa 300
caggggtgtc aagcttccat tttggtatct acttctaaat acactcagaa caggagaaat 360
ttggactaat tttcaaacta cagacacttt ctaatcatga tgcatttcaa aagtggactc 420
gaattaactg agttgcaaaa catgacagtg cccgaggatg ataacattag caatgactcc 480
aatgatttca ccgaagtaga aaatggtcag ataaatagca agtttatttc tgatcgtgaa 540
agtagaagaa gtctcacaaa cagccatttg gaaaaaaaga agtgtgatga gtatattcca 600
ggtacaacct ccttaggcat gtctgttttt aacctaagca acgccattat gggcagtggg 660
attttgggac tcgcctttgc cctggcaaac actggaatcc tacttttttct ggtacttttg 720
acttcagtga cattgctgtc tatatattca ataaacctcc tattgatctg ttcaaaagaa 780
acaggctgca tggtgtatga aaagctgggg gaacaagtct ttggcaccac agggaagttc 840
gtaatctttg gagccacctc tctacagaac actggagcaa tgctgagcta cctcttcatc 900
gtaaaaaatg aactaccctc tgccataaag tttctaatgg gaaaggaaga gacattttca 960
gcctggtacg tggatggccg cgttctggtg gtgatagtta cctttggcat aattctccct 1020
ctgtgtctct tgaagaactt agggtatctt ggctatacta gtggattttc cttgagctgt 1080
atggtttttt tcctaattgt ggttatttac aagaaatttc aaattccctg cattgttcca 1140
gagctaaatt caacaataag tgctaattca acaaatgctg acacgtgtac gccaaaatat 1200
gttaccttca attcaaagac cgtgtatgct ttacccacca ttgcatttgc atttgtttgc 1260
cacccgtcag tcctgccaat ttacagtgag cttaaagacc gatcacagaa aaaaatgcag 1320
atggtttcaa acatctcctt tttcgccatg tttgttatgt acttcttgac tgccatttt 1380

ggctacttga cattctatga caacgtgcag tccgacctcc ttcacaaata tcagagtaaa 1440
gatgacattc tcatcctgac agtgcggctg gctgtcattg ttgctgtgat cctcacagtg 1500
ccggtgttat ttttcacggt tcgttcatct ttatttgaac tggctaagaa aacaaagttt 1560
aatttatgtc gtcataccgt ggttacctgc atactcttgg ttgttatcaa cttgttggtg 1620
atcttcatac cctccatgaa ggatattttt ggagtcgtag gagttacatc tgctaacatg 1680
cttattttca ttcttccttc atctctttat ttaaaaatca cagaccagga tggagataaa 1740
ggaactcaaa gaatttgggc tgccctttc ttgggcctgg gggtgttgtt ctccttggtc 1800
agcattccct tggtcatcta tgactgggcc tgctcatcga gtagtgacga aggccactga 1860
aacccgccga gaaaaagaaa catccctgtt gtctgctcag tcaagtcccc acacatcagc 1920
aatctctcac cacttctttt gcaagtttac agaagcaaac agaaatgtac aggatactta 1980
aaatggaata acttttggt tgcaaaacag agacatggtt ctataatgct tcatgtccct 2040
ccaagatttg agatcaattt agggattgtg aaatttttt tttcaaattt catacaatca 2100
tatttcccag tacttttcac aatcattttt tacccatcta actctatgtt ttgtggcttc 2160
ccggtctctt agaactttga aaacatgata tacaataatg tttatttatt atacatccag 2220
attctgaaat aattttccta ctgatgttca gctcacacta tctgtacctt tttagaagag 2280
aaaagaatct tgaattgtat atatttattt tgctttacag aaaaaaatgg tttcgtaaat 2340
aatttgccta ttttggttaa catagcacat ggagataatc atctgaaagt tatagggcac 2400
tgccactgct gaatcagagc atgcccaata tttgaggtgg ctctgatttc ctggcagctg 2460
aactcgggta gtccagtggc ctagctggta ccacatctat tcccatccag agacattctc 2520
tggcaagtgt tctcagctga aaagtggttg gggatgattc ttaccttggt aattaaatga 2580
agctacacat ttgggtaatc tagcaaatga agtatttttt ccctcttggc aacttgtgtc 2640
agagttactc tggtctgagt caactttcgc tggggaaaac ctatggaacc tactgcaaaa 2700
agattgtcca aaatgcctaa gaaaatactc ctctgatgca tttagccttc aaccctacct 2760
gtcttgctga agggagaaaa atgttttagt acattatagg cccagcagct tttattcatg 2820
tccaccagct agttgcacag agaatcatgt gtacctaact aaggatgatc taggataagt 2880
aactcctgtt ttatattgag tattttaggg aagtctttaa aagacttgtt ttatatctat 2940
aaatctaggt tattacaaat acaagaattt tgtaccttaa ataagcctca tttctatttc 3000
ttcttcatta attctccatc tagtcttgtg aaaaaaaaaa aaaaaaaacc ctcagagata 3060
gtctttgtga agagcttctg acagaatcac tgagtacctt ccttccccca gatgaggaag 3120
acaaggggct ctcagtgtct gtgctgtctc ctcttctctt ccccaaccaa ggactgtgcc 3180
attactgccc gtctcaactg tccatgcagg aggacagagt tgcctggtac tcttaccctt 3240
gtccctctcc taaagggagc acaaggaaac tgaagagact gaaaaagaag agagtttgta 3300
gctgaaaaag aatagggata gcaaggaaac ccagaactgc attcccctaa gtggggccat 3360
cccatgtgat tgaattgtcc atagcttgcc tatggtgaga aatgtgcatg ctccgtgagc 3420
tggtctcttg aaacaggact tatgcttcct ctatattctg gttaaatttt ccaaacacat 3480
aagttcactg agcacagatt tcttatccag agacaagtag aatctaaccg cagactgttg 3540
gcagagtttc caggcactta gccatgttcc cttcctgact caaatcccca aaggccttca 3600
ctctcactga gaatcacact actgtcccat agataaggca ggcattgaag cacctgtcgt 3660
gatcctctag gggggagaat gaaaggttat ttcctgcatt gcatcatcat agcttttaat 3720
ataatgctac agaatcatat ccacattagg ttagagttca gatatttgga tatgaatacc 3780
taacctagcc atatccatgg ccatctctgt tcttttcagc aatgttttcc atattatatt 3840
agcaatgaca gaaacagaac aagccaagat ccagtcagtt cttgggagct tgtctagagc 3900
accaagtaat gaaatagcca ggtagtggga tgactgtacc tttaaaaata cataatttag 3960
tttgcaagct atattatgct actttctatt ttcctcgtta ctttatagca attcatttta 4020
ccctcacaaa gtcaatttag aaccttatca ttaactggga tgtgtagtga tatttttggg 4080
cctctgggtt tcatgtgtca ataccaggca tatctctttc aaatagattt atttagaggg 4140
ggccagtgtt gttgactgtg tggaacccc 4169

<210> 33
<211> 3440
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4027693CB1

<400> 33
gatccccacc acaccaccag cccggccgca cggggcactg agccgggtgc tgagcaccgg 60
aggccccgcc gaggccggga ctcaggacct gcagagaaac gcctcctgat tttgtcttac 120
aatggaactt aaaaagtcgc ctgacggtgg atggggctgg gtgattgtgt ttgtctcctt 180
ccttactcag tttttgtgtt acggatcccc actagctgtt ggagtcctgt acatagaatg 240
gctggatgcc tttggtgaag gaaaaggaaa aacagcctgg gttggatccc tggcaagtgg 300

agttggcttg cttgcaagtc ctgtctgcag tctctgtgtc tcatcttttg gagcaagacc 360
tgtcacaatc ttcagtggct tcatggtggc tggaggcctg atgttgagca gttttgctcc 420
caatatctac tttctgtttt tttcctatgg cattgttgta ggtcttggat gtggtttatt 480
atacactgca acagtgacca ttacgtgcca gtattttgac gatcgccgag gcctagcgct 540
tggcctgatt tcaacaggtt caagcgttgg ccttttcata tatgctgctc tgcagaggat 600
gctggttgag ttctatggac tggatggatg cttgctgatt gtgggtgctt tagctttaaa 660
tatattagcc tgtggcagtc tgatgagacc cctccaatct tctgattgtc ctttgcctaa 720
aaaaatagct ccagaagatc taccagataa atactccatt tacaatgaaa aaggaaagaa 780
tctggaagaa aacataaaca ttcttgacaa gagctacagt agtgaggaaa aatgcaggat 840
cacgttagcc aatggtgact ggaaacaaga cagcctactt cataaaaacc ccacagtgac 900
acacacaaaa gagcctgaaa cgtacaaaaa gaaagttgca gaacagacat atttttgcaa 960
acagcttgcc aagaggaagt ggcagttata taaaaactac tgtggtgaaa ctgtggctct 1020
ttttaaaaac aaagtatttt cagccctttt cattgctatc ttactctttg acatcggagg 1080
gtttccacct tcattactta tggaagatgt agcaagaagt tcaaacgtga aagaagaaga 1140
gtttattatg ccacttattt ccattatagg cattatgaca gcagttggta aactgctttt 1200
agggatactg gctgacttca agtggattaa taccttgtat ctttatgttg ctaccttaat 1260
catcatgggc ctagccttgt gtgcaattcc atttgccaaa agctatgtca cattggcgtt 1320
gctttctggg atcctagggt ttcttactgg taattggtcc atctttccat atgtgaccac 1380
gaagactgtg ggaattgaaa aattagccca tgcctatggg atattaatgt tctttgctgg 1440
acttggaaat agcctaggac cacccatcgt tggttggttt tatgactgga cccagaccta 1500
tgatattgca ttttatttta gtggcttctg cgtcctgctg ggaggtttta ttctgctgct 1560
ggcagccttg ccctcttggg atacatgcaa caagcaactc cccaagccag ctccaacaac 1620
tttcttgtac aaagttgcct ctaatgttta gaagaatatt ggaagacact atttttgcta 1680
ttttatacca tatagcaacg atattttaac agattctcaa gcaaattttc tagagtcaag 1740
actattttct catagcaaaa tttcacaatg actgactctg aatgaattat ttttttttat 1800
atatcctatt ttttatgtag tgtatccgta gcctctatct cgtatttttt tctatttctc 1860
ctccccacac catcaatggg actattctgt tttgctgtta ttcactagtt cttaacattg 1920
taaaaagttt gaccagcctc agaaggcttt ctctgtgtaa agaagtataa tttctctgcc 1980
gactccattt aatccactgc aaggcaccta gagagactgc tcctatttta aaagtgatgc 2040
aagcatcatg ataagatatg tgtgaagccc actaggaaat aaatcattct cttctctatg 2100
tttgacttgc tagtaaacag aagacttcaa gccagccagg aaattaaagt ggcgactaaa 2160
acagccttaa gaattgcagt ggagcaaatt ggtcattttt taaaaaaata tattttaacc 2220
tacagtcacc agttttcatt attctattta cctcactgaa gtactcgcat gttgtttggt 2280
acccactgag caactgtttc agttcctaag gtatttgctg agatgtgggt gaactccaaa 2340
tggagaagta gtcactgtag actttcttca tggttgacca ctccaacctt gctcactttt 2400
gcttcttggc catccactca gctgatgttt cctgggaagt gctaatttta cctgtttcca 2460
aattggaaac acatttctca atcattccgt tctggcaaat gggaaacatc catttgcttt 2520
gggcacagtg gggatgggct gcaagttctt gcatatcctc ccagtgaagc atttatttgc 2580
tactatcaga ttttaccact atcaaatata attcaagggc agaattaaac gtgagtgtgt 2640
gtgtgtgtgt gtgtgtgtgt gtgtgctatg catgctctaa gtctgcatgg gatatgggaa 2700
tggaaaaggg caataagaaa ttaataccct tatgcagttg catttaacct taagaaaaat 2760
gtccttggga taaactccaa tgtttaatac attgattttt tttctaaaga aatgggtttt 2820
aaactttggt atgcatcaga attccctata gatcttttg aaaatatagg tacctgggta 2880
tcacacatag aacttttaat tctgctggtg taggctgttg cccaaacatc tataatttta 2940
ctgagctctt caagtgattc tgataacaca gcctggattg agaatttta taagattggc 3000
aatggaaaaa catttattct tttaaataat aatttttta aaacccaaga ggtcagggga 3060
ttttataaac caatagccaa gtgttcttta aataggaggc acccttccca ttgtgccaaa 3120
atcatctttt catttatttt gaaatttgta tgattatttt atacttgtat gttgcctttc 3180
ttcgaaggcg cctgaagcac tttataaaca caaatcctca caatacctct gtgaggtagg 3240
taaatagtac ttttctatgt agtaaacctg gaatatggag aatttcataa cagttcattc 3300
tacttaataa tgcaataatg gagctccaag ttgtcttgga cttctacacc acactcagac 3360
ttctggaaag ttttctgtac ctcattcttt agtccctgtc aaggttagta aataaaataa 3420
gtgacataaa aaaaaaaaaa 3440

<210> 34
<211> 3699
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472030CB1

<400> 34
atggtttatt ccggaaatgc agagatgttt aacattcaaa aatcaactgc tctaataact 60
gcagaagaac agccaaaact gagaaaggaa gcagttggat ctattgagat attccgcttt 120
gctgatggac tggacatcac actcatgatc ctgggtatac tgacatcact gttcaatgga 180
gcctgccttc ctttaatgcc actgtgtata ggagaaatga gtgataacct tattagtgga 240
tgtctagtcc acactaacac aacaaattat cagaactgta ctcagtctca agagaagctg 300
aatgaagata tgactctgtt gaccctgtat tatgttggaa taggtgttgc tgccttgatt 360
tttggttaca tacagatttc cttgtggatt ataactgcag cacgacagac caagaggatt 420
cgaaaacagt tttttcattc agttttggca caggacatcg gctggtttga tagctgtgac 480
atcggtgaac ttaacactcg catgacagat gacattgaca aaatcagtga tggtattgga 540
gataagattg ctctgttgtt tcaaaacatg tctacttttt cgattggcct ggcagttggt 600
ttggtgaagg gctggaaact caccctagtg actctatcca cgtctcctct tataatggct 660
tcagcggcag catgttctag gatggtcatc tcattgacca gtaaggaatt aagtgcctat 720
tccaaagctg gggctgtggc agaagaagtc ttgtcatcaa tccgaacagt catagccttt 780
agggcccagg agaaagaact tcaaaggtat acacagaatc tcaaagatgc aaaggatttt 840
ggcataaaaa ggactatagc ttcaaaagtg tctcttggtg ctgtgtactt ctttatgaat 900
ggaacctatg gacttgcttt ttggtatgga acctccttga ttcttaatgg agaacctgga 960
tataccatcg ggactgttct tgctgttttc tttagtgtaa tccatagtag ttattgcatt 1020
ggagcagcag tccctcactt tgaaaccttc gcaatagccc gaggagctgc ctttcatatt 1080
ttccaggtta ttgataagaa acccagtata ggtaactttt ccacagctgg atataaacct 1140
gaatccatag aaggaactgt ggaatttaaa aatgtttctt tcaattatcc atcaaggcca 1200
tctatcaaga ttctgaaagg tctgaatctc ggaattaagt ctggagagac agtcgccttg 1260
gtcggtctca atggcagtgg gaagagtacg gtagtccagc ttctgcagag gttatatgat 1320
ccggatgatg gctttatcat ggtggatgag aatgacatca gagctttaaa tgtgcggcat 1380
tatcgagacc atattggagt ggttagtcaa gagcctgttt tgttcgggac caccatcagt 1440
aacaatatca agtatggacg agatgatgtg actgatgaag agatggagag agcagcaagg 1500
gaagcaaatg cgtatgattt tatcatggag tttcctaata aatttaatac attggtaggg 1560
gaaaaaggag ctcaaatgag tggagggcag aaacagagga tcgcaattgc tcgtgcctta 1620
gttcgaaacc ccaagattct gattttagat gaggctacgt ctgccctgga ttcagaaagc 1680
aagtcagctg ttcaagctgc actggagaag gcgagcaaag gtcggactac aatcgtggta 1740
gcacaccgac tttctactat tcgaagtgca gatttgattg tgaccctaaa ggatggaatg 1800
ctggcggaga aaggagcgca tgctgaacta atggcaaaac gaggtctata ttattcactt 1860
gtgatgtcac aggatattaa aaaagctgat gaacagatgg agtcaatgac atattctact 1920
gaaagaaaga ccaactcact tcctctgcac tctgtgaaga gcatcaagtc agacttcatt 1980
gacaaggctg aggaatccac ccaatctaaa gagataagtc ttcctgaagt ctctctatta 2040
aaaattttaa agttaaacaa gcctgaatgg ccttttgtgg ttctggggac attggcttct 2100
gttctaaatg gaactgttca tccagtattt tccatcatct ttgcaaaaat tataaccatg 2160
tttggaaata atgataaaac cacattaaag catgatgcag aaatttattc catgatattc 2220
gtcattttgg gtgttatttg ctttgtcagt tatttcatgc aggatattgc ctggtttgat 2280
gaaaaggaaa acagcacagg aggcttgaca acaatattag ccatagatat agcacaaatt 2340
caaggagcaa caggttccag gattggcgtc ttaacacaaa atgcaactaa catgggactt 2400
tcagttatca tttcctttat atatggatgg gagatgacat tcctgattct gagtattgct 2460
ccagtacttg ccgtgacagg aatgattgaa accgcagcaa tgactggatt tgccaacaaa 2520
gataagcaag aacttaagca tgctggaaag atagcaactg aagctttgga gaatatacgt 2580
actatagtgt cattaacaag ggaaaaagcc ttcgagcaaa tgtatgaaga gatgcttcag 2640
actcaacaca gaaatacctc gaagaaagca cagattattg gaagctgtta tgcattcagc 2700
catgccttta tatattttgc ctatgcggca gggtttcgat ttggagccta tttaattcaa 2760
gctggacgaa tgaccccaga gggcatgttc atagttttta ctgcaattgc atatggagct 2820
atggccatcg gagaaacgct cgttttggct cctgaatatt ccaaagccaa atcgggggct 2880
gcgcatctgt ttgccttgtt ggaaaagaaa ccaaatatag acagccgcag tcaagaaggg 2940
aaaaagccag acacatgtga agggaattta gagtttcgag aagtctcttt cttctatcca 3000
tgtcgcccag atgttttcat cctccgtggc ttatccctca gtattgagcg aggaaagaca 3060
gtagcatttg tggggagcag cggctgtggg aaaagcactt ctgttcaact tctgcagaga 3120
ctttatgacc ccgtgcaagg acaagtgctg tttgatggtg tggatgcaaa agaattgaat 3180
gtacagtggc tccgttccca aatagcaatc gttcctcaag agcctgtgct cttcaactgc 3240
agcattgctg agaacatcgc ctatggtgac aacagccgtg tggtgccatt agatgagatc 3300
aaagaagccg caaatgcagc aaatatccat tcttttattg aaggtctccc tgagaaatac 3360
aacacacaag ttggactgaa aggagcacag ctttctggcg gccagaaaca aagactagct 3420
attgcaaggg ctcttctcca aaaacccaaa attttattgt tggatgaggc cacttcagcc 3480
ctcgataatg acagtgagaa ggtggttcag catgcccttg ataaagccag gacgggaagg 3540
acatgcctag tggtcactca caggctctct gcaattcaga acgcagattt gatagtggtt 3600
ctgcacaatg gaaagataaa ggaacaagga actcatcaag agctcctgag aaatcgagac 3660
atatatttta agttagtgaa tgcacagtca gtgcagtga 3699

<210> 35
<211> 2428
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476089CB1

<400> 35
atagatcctg aaaaggaaac gactgatatc accatcaagc tagtgatcat ccatatggct 60
tgctgcagtt ctccacaggg ctgcctcctc agcctaagga cgcatgaccc tgcctgcaag 120
cagcgttcca catatcactg tggaggagga agatggagaa atcaggttat ggtcatccgt 180
gcacacggga cttctgggaa gggtgactgc ggaatttaga acagtgtcct tgacagcatt 240
cagtcctgag gattaccaga atgttgctgg cacattagaa tttcaaccag gagaaagata 300
taaatacatt ttcataaaca tcactgataa ttctattcct gaactggaaa aatcttttaa 360
agttgagttg ttaaacttgg aaggaggagc cagtctagga gtggcttccc aaattctagt 420
gacaattgca gcctctgacc acgctcatgg cgtatttgaa tttagccctg agtcactctt 480
tgtcagtgga actgaaccag aagatgggta tagcactgtt acattaaatg ttataagaca 540
tcatggaact ctgtctccag tgactttgca ttggaacata gactctgatc ctgatggtga 600
tctcgccttc acctctggca acatcacatt tgagattggg cagacgagcg ccaatatcac 660
tgtggagata ttgcctgacg aagacccaga actggataag gcattctctg tgtcagtcct 720
cagtgtttcc agtggttctt tgggagctca tattaatgcc acgttaacag ttttggctag 780
tgatgatcca tatgggatat tcatttttc tgagaaaaac agacctgtta aagttgagga 840
agcaacccag aacatcacac tatcaataat aaggttgaaa ggcctcatgg gaaaagtcct 900
tgtctcatat gcaacactag atgatatgga aaaaccacct tattttccac ctaatttagc 960
gagagcaact caaggaagag actatatacc agcttctgga tttgctcttt ttggagctaa 1020
tcagagtgag gcaacaatag ctatttcaat tttggatgat gatgagccag aaaggtccga 1080
atctgtcttt atcgaactac tcaactctac tttagtagcg aaagtacaga gtcgttcaat 1140
tccaaattct ccacgtcttg ggcctaaggt agaaactatt gcgcaactaa ttatcattgc 1200
caatgatgat gcatttggaa ctcttcagct ctcagcacca attgtccgag tggcagaaaa 1260
tcatgttgga cccattatca atgtgactag aacaggagga gcatttgcag atgtctctgt 1320
gaagtttaaa gctgtgccaa taactgcaat agctggtgaa gattatagta tagcttcatc 1380
agatgtggtc ttgctagaag gggaaaccag taaagccgtg ccaatatatg tcattaatga 1440
tatctatcct gaactggaag aatcttttct tgtgcaactg atgaatgaaa caacaggagg 1500
agccagacta ggggctttaa cagaggcagt cattattatt gaggcctctg atgaccccta 1560
tggattattt ggttttcaga ttactaaact tattgtagag gaacctgagt ttaactcagt 1620
gaaggtaaac ctgccaataa ttcgaaattc tgggacactc ggcaatgtta ctgttcagtg 1680
ggttgccacc attaatggac agcttgctac tggcgacctg cgagttgtct caggtaatgt 1740
gacctttgcc cctggggaaa ccattcaaac cttgttgtta gaggtcctgg ctgacgacgt 1800
tccggagatt gaagaggtta tccaagtgca actaactgat gcctctggtg gaggtactat 1860
tgggttagat cgaattgcaa atattattat tcctgccaat gatgatcctt atggtacagt 1920
agcctttgct cagatggttt atcgtgttca agagcctctg gaaagaagtt cctgtgctaa 1980
tataactgtc aggcgaagcg gagggcactt tggtcggctg ttgttgttct acagtacttc 2040
cgacattgat gtagtggctc tggcaatgga agaaggtcaa gatttactgt cctactatga 2100
atctccaatt caaggggtgc ctgacccact ttggagaact tggatgaatg tctctgccgt 2160
gggggagccc ctgtatacct gtgccacttt gtgccttaag gaacaagctt gctcagcgtt 2220
ttcatttttc agtgcttctg agggtcccca gtgtttctgg atgacatcat ggatcagccc 2280
agctgtcaac aattcagact tctggaccta caggaaaaac atgaccaggg tagcatctct 2340
tttagtggtc aggctgtggc tgggagtgac tatgagcctg tgacaaggca atgggccata 2400
atgcaggaag gtgatgaatt cgcaaaaa 2428

<210> 36
<211> 2243
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6428177CB1

<400> 36
gtaactccag gacgagaccg gagcgacccg cgcagagcat aggcggcgaa ctgcgcccgg 60

cgcccgagac cggcagctgc gtggggcggg ggctgcgccc gagcccgatc tgccggctcc 120
gagtggtctc ggaaagaggg tcgtggtccc gcacggatgc gcttgttggg agaaaccttg 180
gagattcacg gcaaggcgta aagcctgggg cttccaacga tactctgggc agggatggaa 240
gcctagatgc ctcaccgcaa ggagcggccg agcgggtcct cgcttcacac acacggcagc 300
accggcaccg cggagggagg aaacatgtcc cggctgtctc tcacccggtc gcctgtgtct 360
cccctggctg cccagggcat ccccctgcca gcccagctca ccaagtccaa tgcacctgtg 420
cacatcgatg tgggcggcca catgtacacc agcagcctgg ccacgctcac caagtaccct 480
gactccagga taagccgcct cttcaatggc actgaaccca tcgtcctgga cagtttgaag 540
caacattatt tcattgaccg ggatggggag attttccgct acgtcctgag cttcctgcgg 600
acgtccaagc tgctgcttcc ggatgacttt aaggacttca gtctgctgta cgaggaggcg 660
cgctactatc agctccagcc catggtgcgc gagctggagc gctggcagca ggagcaggag 720
cagcggcgcc gcagccgggc ctgtgactgc ctggtggtgc gcgtcacgcc cgacttgggc 780
gagcggatcg cactcagcgg cgagaaggcc ctcatcgagg aggtcttccc cgagaccgga 840
gacgtcatgt gcaactccgt caacgccggc tggaaccagg accccacgca cgtcatccgc 900
ttcccgctca atggctactg ccggctcaac tcggtacagg tcctggagcg gctgttccag 960
aggggtttca gcgtggctgc gtcctgtggg ggcggtgtgg actcctccca gttcagcgag 1020
tatgtgcttt gccgggagga gcggcggccg cagcccaccc ccactgctgt tcgaatcaag 1080
caggaacccc tggactaggc cctgcttcag tgcccacctg ggcccccca gggacctgga 1140
aacagtgctg gggagttctg cctgtgtata cttggccgtg ggcatgagac cgagggtgag 1200
gctggagggt ccaaagctgg cccagcgagc accagggtcc caggtgtcat ggcaacagaa 1260
cgtgggatgc tggaggcatg cctgcagaag gactgttgat gcgacccaaa gatacagcgg 1320
tgggatctct gctgccagct ctcccagccc ctcagcttcg cagcctggcg cagcatcctc 1380
tgaggccccg gggcctgttg gggcggggtt ggaagagccg tctgcagcta cttcagagga 1440
gctgtttatc cctctccacg cggggcagac tctggcgggt ctcctagcgt ccgagagatg 1500
gcttattttc tacagtattt aaaatggatg cagccctaac tgcaaaagtc agagaggctg 1560
acaaggacca atgcttcttt atctggtgct cagttctcag tcagacgtgc agcatggctg 1620
cagggtggac cagctgcctg gcattcaggc ccagatgcct gcagggctgg ggctctcggg 1680
acagatgcag ggatgtgtgc tgcagggctg ctgggaggag agtggtgggg gcctgagggc 1740
tgagtgattc tgtaaccacc tgagaccttc acgtttgctg ccgttggggg ctcaggctgc 1800
actccccggg tcacctgacc tgctgcccag gggcttccag tcctgtctgt gtggactggc 1860
acctgggctg ctggagaagt ctcctcccgt tcggaccagc ctcagggctg cacgttacct 1920
caggaatggg ccccaccatg aaggggccca tctgtcagca gcgtcttcta ggtccccagc 1980
tcagggagcc atccccagct ccagttttct catgcgaata tgcacagttt taattcacgt 2040
tgttacacta gcctgccgat gagacccaga cacaggcaga cctggcgctc ttgacccctg 2100
attccagtga ggactggccc tgaggagtcc ttgcagacct gctgcctgcc ccacgacagg 2160
cccaaagatg gacccccct ggccttgtga cagctcccca agtgttctcc ggtggagaaa 2220
ctgcagagga ctggtgggcg ggg 2243

<210> 37
<211> 3711
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477243CB1

<400> 37
gagcgggtgg cccggccgcc cgcctcgctg ctccgcttgg cgccgccggc ccacgccgca 60
gtgtgttttg tggacggcgc cttcccagac agcccggtag agcccagctc agcgcccggc 120
agccttcgac gcgatgttcc gccggagctt gaatcgtttt tgtgctggag aagagaaacg 180
agttggcaca cgcacagtgt ttgttggcaa tcatccagtt tcggaaacag aagcttacat 240
tgcacaaaga ttttgtgata atagaatagt ctcatctaag tatacacttt ggaatttttct 300
cccaaagaat ctgtttgaac agtttagaag aattgcaaat ttttattttc tcataatctt 360
ccttgtacag gtcacagtag acacaccaac tagcccagtt accagtggac ttccactttt 420
ctttgttata actgttacag ccatcaagca gggatatgag gattgtctga gacacagagc 480
tgacaatgaa gtcaacaaaa gcactgttta cattattgaa aatgcaaagc gagtgagaaa 540
agaaagtgaa aaaatcaagg ttggtgatgt agtagaagta caggcagatg aaacctttcc 600
ctgtgatctt attcttctat catcttgcac cactgatgga acctgttatg tcactacagc 660
cagtcttgat ggggaatcca attgcaagac acattatgca gtacgtgata ccattgcact 720
gtgtacagca gaatccatcg ataccctccg agcagcaatt gaatgtgaac agcctcaacc 780
tgacctctac aaatttgttg ggcgaatcaa tatctacagt aatagtcttg aggctgttgc 840
caggtctttg ggacctgaaa atctcttgct gaaaggagct acgctaaaaa ataccgagaa 900

gatatatgga gttgctgttt acactggaat ggaaaccaaa atggctttga actaccaagg 960
gaaatctcag aaacgttctg ctgttgaaaa atctattaat gctttcctga ttgtatattt 1020
atttatctta ctgaccaaag ctgcagtatg cactactcta aagtatgttt ggcaaagtac 1080
cccatacaat gatgaacctt ggtataacca aaagactcag aaagagcgag agaccttgaa 1140
ggttttaaaa atgttcaccg acttcctatc atttatggtt ctattcaact ttatcattcc 1200
tgtctccatg tacgtcacag tagaaatgca gaaattcttg ggctccttct tcatctcatg 1260
ggataaggac ttttatgatg aagaaattaa tgaaggagcc ctggttaaca catcagacct 1320
taatgaagaa cttggtcagg tggattatgt atttacagat aagactggaa cactcactga 1380
aaacagcatg gaattcattg aatgctgcat agatggccac aaatataaag gtgtaactca 1440
agaggttgat ggattatctc aaactgatgg aactttaaca tattttgaca aagtagataa 1500
gaatcgagaa gagctgtttc tacgtgcctt gtgtttatgt catactgtag aaatcaaaac 1560
aaacgatgct gttgatggag ctacagaatc agctgaatta acctatatct cctcttcacc 1620
agatgaaata gctttggtga aaggagctaa aaggtacggg ttcacatttt taggaaatcg 1680
aaatggatat atgagagtag agaaccaaag aaaagaaata gaagaatatg aacttcttca 1740
caccttaaac tttgatgctg tccggcgacg tatgagtgta attgtgaaga ctcaagaagg 1800
agacatactt ctcttttgta aaggagcaga ctcggcagtt tttcccagag tgcaaaatca 1860
tgaaattgag ttaactaaag tccatgtgga acgtaatgca atggatgggt atcggacact 1920
ctgtgtagcc ttcaaagaaa ttgctccaga tgattatgaa agaattaaca gacagctcat 1980
agaggcaaaa atggccttac aagacagaga agaaaaaatg gaaaaagttt tcgatgatat 2040
tgagacaaac atgaatttaa ttggagccac tgcagttgaa gacaagctac aagatcaagc 2100
tgcagagacc attgaagctc tgcatgcagc aggcctgaaa gtctgggtgc tcactgggga 2160
caagatggag acagctaaat ccacatgcta tgcctgccgc cttttccaga-ccaacactga 2220
gctcttagaa ctaaccacaa aaaccattga agaaagtgaa aggaaagaag atcgattaca 2280
tgaattattg atagaatatc gcaagaaatt gctgcatgag tttcctaaaa gtactagaag 2340
ctttaaaaaa gcatggacag aacatcagga atatggatta atcatagatg gctccacatt 2400
gtcactcata ctaaattcta gtcaagactc tagttcaaac aattacaaaa gcattttcct 2460
acaaatatgt atgaagtgta ctgcagtgct ctgctgtcgg atggcaccat tacagaaagc 2520
ccagattgtc agaatggtga agaatttaaa aggcagccca ataactctgt cgataggtga 2580
tggtgccaat gatgttagta tgatcttgga atcccatgtg ggaataggta ttaaaggcaa 2640
agaaggtcgc caagcagcta ggaatagcga ttattctgtt ccaaagttta aacacttaaa 2700
gaaactgctg ttggctcatg gacatctata ttatgtgaga atagcacacc ttgtacagta 2760
cttcttctat aagaaccttt gtttcatttt gccacagttt ttgtaccagt tcttctgtgg 2820
attctcacaa cagccactgt atgatgctgc ttaccttaca atgtacaata tctgcttcac 2880
atccttgccc atcctggcct atagtctact ggaacagcac atcaacattg acactctgac 2940
ctcagatccc cgattgtata tgaaaatttc tggcaatgcc atgctacagt tgggcccctt 3000
cttatattgg acatttctgg ctgcctttga agggacagtg ttcttctttg ggacttactt 3060
tctttttcag actgcatccc tagaagaaaa tggaaaggta tacggaaact ggacttttgg 3120
aaccattgtt tttacagtct tagtattcac tgtaaccctg aagcttgcct tggatacccg 3180
attctggacg tggataaatc actttgtgat ttggggttct ttagccttct atgtattttt 3240
ctcattcttc tggggaggaa ttatttggcc ttttctcaag caacagagaa tgtatttttgt 3300
atttgcccaa atgctgtctt ctgtatccac atggttggct ataattcttc taatatttat 3360
cagcctgttc cctgagattc ttctgatagt attaaagaat gtaagaagaa gaagtgccag 3420
gagaaatctg agctgtagaa gggcatctga ctcattatcc gccagacctt cagtcagacc 3480
tcttctttta cgaacattct cagacgaatc taatgtattg taacagaatc cgaatcttga 3540
actgcctatg ttattgtcct acaagcatac tgacagtggt tacagctaaa aaagaaagca 3600
tgaagaaaca actacaaaaa gttatcatct caggatactt gatatgcaac acactaaacc 3660
actctcatgt ctagaatcac aataaatttc attaattgag ggtagaggtt a 3711

<210> 38
<211> 2717
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473042CB1

<400> 38
cccgccgggc tccaactccg cagcgtcgga gcgcggcggg cagcaacttt ctccccggag 60
cggccgtggc ggcggctgct gccgtggcag ccggagcgga agccgggagg aagaaagcgg 120
cggcagcggc ggttgctccc gccggctcgg gctgtctagc tcgccgagac tgccggcccg 180
cggagccgcg taccccgggg cagccccggg cccctgccct atgtcccgca aggcaagcga 240
gaatgtggag tacacgctgc ggagcctgag cagcctgatg ggcgagcggc gcaggaagca 300

gccggagccg gacgcggcga gcgcggccgg ggagtgcagc ctcctggctg ccgccgaatc 360
gagcaccagc ctgcagagcg cgggcgcggg cggcggcggc gtcggggacc tggagcgcgc 420
ggcgcggcgg cagttccagc aggacgagac ccccgccttc gtgtacgtgg tggccgtctt 480
ctccgcgctg ggcggcttcc tgtttggcta tgacaccggg gtggtgtcag gggccatgct 540
gctgctcaag cggcagctca gtctggacgc gctgtggcag gagctgctgg tgtccagcac 600
ggtgggggcg gctgccgtct cggcgctggc cggaggcgcc ctcaacggcg tcttcggccg 660
ccgcgctgcc atcctcctgg ccagtgccct cttcaccgcc ggctccgcgg tgctggctgc 720
ggccaacaac aaggagacac tgctcgccgg ccgcctggtc gtgggactcg gcatcggcat 780
tgcttctatg acagtgccag tgtacattgc ggaggtctca ccacccaatt taagaggccg 840
attagtcacc attaataccc tcttcatcac aggagggcag ttctttgcaa gtgttgttga 900
tggagccttc agttatctcc agaaggatgg atggaggtac atgttgggac ttgcagtagt 960
tccggcggtt atacagtttt ttggctttct ctttttgcct gaaagccctc gatggcttat 1020
tcagaaagga cagactcaga aggcccgtag aattttatct cagatgcgtg gtaaccagac 1080
cattgatgag gaatatgata gcatcaaaaa caacattgaa gaggaggaaa aagaggttgg 1140
ctcagctgga cctgtgatct gcagaatgct gagttatccc caaactcgcc gagctttaat 1200
tgtgggttgt ggcctacaaa tgttccagca gctctcaggc attaacacca tcatgtacta 1260
cagtgcaacc attctgcaga tgtctggtgt tgaagatgat agacttgcaa tatggctggc 1320
ttcagttaca gccttcacaa atttcatttt cacacttgtg ggagtctggc ttgttgagaa 1380
ggtgggccgc agaaagctta cctttggtag tttagcaggt accaccgtag cactcattat 1440
tcttgccttg ggatttgtgc tatcagccca agtttcccca cgcatcactt ttaagccaat 1500
agctccgtca ggtcagaacg ccacttgcac aagatacagt tactgtaatg aatgtatgtt 1560
ggatccagac tgcggtttct gctacaagat gaacaaatca actgtcattg actcctcctg 1620
tgttccagtt aataaagcat ctacaaatga ggcagcctgg ggcaggtgtg aaaatgaaac 1680
caagttcaaa acagaagata tattttgggc ttacaatttc tgccctactc catactcctg 1740
gactgcactt ctgggcctta ttttatatct tgtcttcttt gcacctggaa tgggaccaat 1800
gccttggact gtgaattctg aaatatatcc cctttgggca agaagtacag gaaatgcatg 1860
ttcatctgga ataaactgga ttttcaatgt cctggtttca ctaacatttt tacacacagc 1920
agagtatctt acatactatg gagctttctt cctctatgct ggatttgctg ctgtgggact 1980
ccttttcatc tatggctgtc ttcctgagac caaaggcaaa aaattagagg aaattgaatc 2040
actctttgac aacaggctat gtacatgtgg cacttcagat tctgatgaag ggagatatat 2100
tgaatatatt cgggtaaagg gaagtaacta tcatctttct gacaatgatg cttctgatgt 2160
ggaataattt tcagctgctc atatatttag ttatttaaac aaactggggg gagaagaaca 2220
gcaattggtg acttcactgc cctgcttcta atctggttct ttccacagcc tagttttgat 2280
tgacttcata ttctagaata cttgattagg aggaagatac aaccatgatg actttttttt 2340
tccacaagga acaatatttt aaaaaatatt tacagagatt ttaatctaat aattcttaag 2400
caaatgtgtg taatgccttc ctgaaatagt ctaaaatgaa tattgtaccc agtgacttca 2460
gtggtatcct tttttcctaa gaccatttat aattattagt ggcaacagag tcagtgctaa 2520
tctagccaaa ttacatatgt ataatatatt tataaaggat tctgggagat ggtccaaggg 2580
tgttctgtgt caaaagatgg cctattggcc ctcagttttc ctacagagta gtggcttatc 2640
tctgatcagc tgttacaaac taaattccat gtaagctttc atcaacaaat tccaaagtgc 2700
ctcctacaag ggcacag 2717

<210> 39
<211> 2235
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482060CB1

<400> 39
agggagcgcc ggagacgggg agctattccg ccccggcggc tccattcggc gcccgcagcc 60
ctcagggggt cggccccgcg gcttgggaga gggcaccgcg gcctcggtgt gcgcagccct 120
cgggcgcgag ggtcggcggc gcggacacag ccgcgttccc agccggtggg gctcagcgct 180
ggcgccggcg aggactcccc ggccacccgc aggtaccgcc gggcggaggg cgcgctacta 240
gcagcgccgg agatactcga gcccagggac ccccgggcca gcggagggca ggagcggagc 300
cccgagggag cgcgggcccc gacggcgcgc tccccccgtca gccacgggca ggcaggcccc 360
gcgtggcggc ttggggtggg gggctgcagc ggggccctcg ggccgaaagt cccccgggcg 420
gccagccatg accttcgggc gcagcggggc ggcctcggtg gtgctgaacg tgggcggcgc 480
ccggtattcg ctgtcccggg agctgctgaa ggacttcccg ctgcgccgcg tgagccggct 540
gcacggctgc cgctccgagc gcgacgtgct cgaggtgtgc gacgactacg accgcgagcg 600
caacgagtac ttcttcgacc ggcactcgga ggccttcggc ttcatcctgc tctacgtgcg 660

cggccacggc aagctgcgct tcgcgccgcg gatgtgcgag ctctccttct acaacgagat 720
gatctactgg ggcctggagg gcgcgcacct cgagtactgc tgccagcgcc gcctcgacga 780
ccgcatgtcc gacacctaca ccttctactc ggccgacgag ccgggcgtgc tgggccgcga 840
cgaggcgcgc cccggcgcgc gaggcggctc cctccaggcg ctggctggag cgcatgcggc 900
ggaccttcga ggagcccaca tcctggctag cgtgtcggtg gtgttcgtga tcgtgtccat 960
ggtggtgctg tgcgccagca cgttgcccga ctggcgcaac gcagccgccg acaaccgcag 1020
cctggatgac cggagcagga taattgaagc tatctgcata ggttggttca ctgccgagtg 1080
catcgtgagg ttcattgtct ccaaaaacaa gtgtgagttt gtcaagagac ccctgaacat 1140
cattgattta ctggcaatca cgccgtatta catctctgtg ttgatgacag tgtttacagg 1200
cgagaactct caactccaga gggctggagt caccttgagg gtacttagaa tgatgaggat 1260
tttttgggtg attaagcttg cccgtcactt cattggtctt cagacactcg gtttgactct 1320
caaacgttgc taccgagaga tggttatgtt acttgtcttc atttgtgttg ccatggcaat 1380
ctttagtgca ctttctcagc ttcttgaaca tgggctggac ctggaaacat ccaacaagga 1440
ctttaccagc attcctgctg cctgctggtg ggtgattatc tctatgacta cagttggcta 1500
tggagatatg tatcctatca cagtgcctgg aagaattctt ggaggagttt gtgttgtcag 1560
tggaattgtt ctattggcat tacctatcac ttttatctac catagctttg tgcagtgtta 1620
tcatgagctc aagtttagat ctgctaggag catttgccta acaagtgtca cttctgtgct 1680
gggcactgtg gggtatacag agatgaccat caacgggcct tgccctgacg ccctgagaga 1740
tccttgtacc tgcaaaaagc ccttgaagac ccattctggg gtcctttaca aggccatggc 1800
tgatttgtgg cagtctctag aaggtggccc accggtggag cagctgcccc cagacccctt 1860
gacccggtgg tgcttccacc ctgccggaag caccttgtgt ggccccgcca acagcatggc 1920
ggttgcatcc ccaggaagca ggcccgcagc gcccggaggg ggtttcctga ggacagaggc 1980
ccttgtcctg attgtcgcag caggccctgt cgatggactt aactgtgaaa atcacccttt 2040
caggggtgga tgcaaggact tctgagggcg gagaagtaga taccttcctg gatagctgtg 2100
gagccggggt cctgcatttc cctctggcgt ctctgctgac tgagatgtga agcagtcggc 2160
ccatgtcccg agaagaggtt ggccacaact ctgtgccaca tgctcttcat tttagaatcc 2220
aggatgaagg atatg 2235

<210> 40
<211> 2563
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1578772CB1

<220>
<221> unsure
<222> 2466
<223> a, t, c, g, or other

<400> 40
gagatggaag ataacaatgt tccagtaaga cctttgtagg ctctaagcgc tgaaaaagtt 60
tatggtgccc caggtgtggt tcaaaataaa aacacaaaac tataaaataa acttaagaga 120
taaaattctt attttttca cacaatgaca ttgtctttta aaaggaatat cacatatcaa 180
agtctccccc taatattttt ggtttccctt actttgctgg tgccctaaac acatataccc 240
tgcttatagg gtaacccaac agtggctttt aaggtggagg tgggctacaa aactggggaa 300
cttgaacaat ggacatctac aaggagactg gtgaataatg ggcatttaat taattggggg 360
gagcgggagg gttagatgtg aatggcagaa tattaagaag ggggttgtgt ggaaggagat 420
ttgggagaag ggagacttcc gaggaagatt aggcagagtg ggcaggaaga ccagctctca 480
tgtgggggtg ggaggctctc ttccttttt gctcctgttc ctccttttct ctagctggca 540
ggcctcttct cctgcacagt ggtcctgtcg gtgctgctgt ggctggggcc cttcttttac 600
tatctgccca aggctgtcct ggcttgcatc aacatctcca gcatgcgcca ggtgttctgc 660
cagatgcagg aacttccaca actatggcac atcagccgag tggactttgc tgtgtggatg 720
gtcacctggg tggcagtagt gaccctgagt gtggatttgg gcctggctgt gggtgtggtc 780
ttctccatga tgactgtggt ctgccgcacc cggagctcct ccaggtcccg gggctctgca 840
tcctgagcta tccaacacca ctgtactttg ggacccgtgg gcagtttcgc tgcaacctgg 900
agtggcacct ggggctcgga gaaggagaaa aggagacttc aaagccagat ggcccaatgg 960
ttgcagttgc tgagcctgtc agggtggtgg tcctagactt cagtggtgtc acctttgcag 1020
atgctgctgg ggccagagaa gtggtgcagc tggccagccg atgtcgagat gctaggatcc 1080
gcctcctcct ggctcagtgt aatgccttgg tgcaggggac actgacccgg gtaggactcc 1140
tggacagggt gactccagat cagctgtttg tgagtgtgca ggatgcagct gcttatgccc 1200

```
tggggagcct ggtaaggggc agtagcacca ggagcgggag ccaggaggca ctgggctgcg 1260
gcaagtgagg caggggagct cactgaccca aagatttgca ccgtgtgggt ctgacctcat 1320
catgtggagt gcagagggcc ctgatgacat gtgtgtgatg aggaccatga cccttgaacc 1380
cccttaccta acgtaactaa taaaatgaag ctgagagctt tggaatccat gaagtgagtc 1440
taggtgtttg cacagggact ctggtgcccc ttcttttgtg cccacagcat tgcagagaca 1500
caactaagaa tggctttcac caaccaccag ccctcacccc agccccagag ccacaagttc 1560
tctctgtggg ggtggggctg gagcaggtac acagagtact ggatctgaag atgcagatga 1620
ggggcacagt cttgggatta tgtgttgggg aacttcccca ccccctcggt cccaagatga 1680
gaggacagtg tttccacctt aggttcttag agtccctctg ggctctttgg cacttggaaa 1740
gtgatccccc catttcctgc cccatgagaa tgggcagggg gaggacttgg cactggctgt 1800
gggagaggtt atggctccac caggcctctg ggcactggaa aaaggagggg tgtcaccagg 1860
acaaccccta ctgggcatca ggttctgaaa aggaagagtg aggaactaga ggctcaggga 1920
cagccagtag agtgctcaca gggtggtggg gtttgttgga aattcctggc agggacaagg 1980
aggcaggccc agcctgacag tcagaaatcc ccagcgggcc atcactggga ggtcatgcac 2040
tgcagccggt gtttgaagaa gagcagccgg tgtttggcca tctggctctc gtccagtgat 2100
cctgggtaac ggtaccgggc gtaagtctct gctccggcat cccttgatgt ccaaggcagc 2160
ttcagtttgg atgcatgatc caccacgacg tcggagcagg agccaacccg aaggtaacca 2220
agcccatcca agaagaattc cagatgagcc acgcggctga agcgggggtc gaaaccgacc 2280
tcgcgcacct tgtcagtccg cgccaggaag aagttaacca cgccgtcggt gaccacgcag 2340
cctgggaagc cgacgagctc gtggtggaag cagcgccttt gccggaggag ttcccgaggc 2400
ctggggcgcc gggctccacg ctcagcagct gccgataagt ggtggcaaag ccggagatct 2460
cgcgcnacgg ccccccacca ggtccagcgg cgtccgtcca gcacgtcaca agcttctcag 2520
cgcgtccggc cgtgaagacg aatcgttcgt caccacagca cga                   2563
```